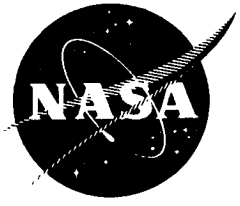
PM-81

THIRD COMPILATION

OF PAPERS ON

TRAJECTORY ANALYSIS

AND GUIDANCE THEORY



ELECTRONICS RESEARCH CENTER
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

PM-81

THIRD COMPILATION OF PAPERS ON TRAJECTORY ANALYSIS AND GUIDANCE THEORY

AUGUST 1969

Prepared by Contractors
for the Computer Research Laboratory
NASA Electronics Research Center
Cambridge, Massachusetts

FOREWORD

This volume contains 12 papers prepared by agencies working in trajectory analysis and guidance theory with the Computer Research Laboratory of the NASA Electronics Research Center. The papers are concerned with special studies performed in guidance theory, optimization theory, numerical methods, and celestial mechanics. They include:

1. An extension of the classical theory of calculus of variations to include varying number and types of constraints;

2. A development of theory for relaxed controls for integral equations;

3. A generalization of the above case to one where the class of controls may, but need not, consist of relaxed controls;

4. An application of Hamilton-Jacobi theory to a planar trajectory optimization problem;

5. A presentation of a method of obtaining a complete integral of the Hamilton-Jacobi equation associated with a dynamical system in which constants of motion are known;

6. A method of solving two-point boundary-value problems by an offset vector iteration method;

7. A linearized guidance procedure based on minimum impulses for space trajectories;

8. A set of equations for computing orbits in closed form using the spheroidal method of calculation; in particular, they are good for polar and near-polar orbits;

9. A procedure for developing expansion formulas in canonical transformation in which the form is developed for speedy computerized symbolic manipulation;

10. A formal solution of the n-body problem in Taylor series;

11. A paper on the long period behavior of a close lunar orbiter;

FOREWORD

12. A presentation of non-linear resonance theory with an application.

These papers cover work performed from 1 October 1967 to 1 February 1969. This work was supervised by personnel of the Computer Research Laboratory.

SUMMARY

This volume contains technical papers on NASA-sponsored studies in the areas of trajectory analysis and guidance theory. These papers cover the period beginning 1 October 1967 and ending 1 October 1968. The technical supervision of this work is under the personnel of the Computer Research Laboratory at NASA-ERC.

CONTENTS

CONTENTS (Concl'd)

INTRODUCTION

By William E. Miner
Chief, Computation Theory
and Techniques Branch
NASA Electronics Research Center

This document contains 12 technical papers covering work sponsored by the Computer Research Laboratory of the NASA Electronics Research Center in the fields of guidance theory, optimization theory, numerical methods, and celestial mechanics.

The following table lists the authors, contributing institutions, and the disciplines of each paper.

Author	Institution/Company	Discipline
J. L. Linnstaedter	Arkansas State Univ.	Optimization Theory
J. Warga*	Northeastern Univ.	Optimization Theory
S. K. Lakhanpal	Vanderbilt Univ.	Optimization Theory
P. M. Fitzpatrick/ J. E. Cochran	Auburn Univ.	Optimization Theory
C. F. Price	MIT	Numerical Methods
T. N. Edelbaum	AMA	Guidance Theory
J. P. Vinti	MIT	Celestial Mechanics
A. A. Kamel	Stanford Univ.	Celestial Mechanics
P. Sconzo/ D. Valenzuela	IBM	Celestial Mechanics
R. Dasenbrock	Stanford Univ.	Celestial Mechanics
W. T. Kyner	Univ. of So. Calif.	Celestial Mechanics

*Two papers

The above characterization is made only in a general way. Work done in optimization theory may have application in trajectory analysis, control theory, guidance theory, and/or celestial mechanics. Work done in celestial mechanics often overlaps into the area of optimization theory with potential applications to that theory. Numerical methods find usages in many disciplines.

INTRODUCTION

Synopses of the individual papers are presented below:

Paper No. 1

The first paper, written by J. L. Linnstaedter of Arkansas State University, presents a generalized, multistage problem of Bolza in the calculus of variations. The differential constraints and the number of differential constraints may be different for each of the various stages. The stages are allowed to degenerate. Discontinuities at staging points are permitted. The paper presents a multiplier rule and analogues of the Weierstrass and Clebsch conditions.

Paper No. 2

The second paper, written by J. Warga of Northeastern University, covers relaxed controls for functional equations where the functional values of the state are considered as known functions of states and controls. These equations are constrained by known functions of the states and controls, and a function of the controls and states is minimized. The controls are embedded in a set of "relaxed controls" so that the existence of a relaxed minimizing point and an approximate solution may be obtained under mild assumptions. Theorem 2.1 presents the results described above. The proof is presented in paragraph 5. The paper presents theorems based on the special case of a control problem defined by a Uryson-type integral equation.

Paper No. 3

The third paper, also written by J. Warga of Northeastern University, is a generalization of the second paper in this compilation. The existence of an original (unrelaxed) control is assumed. It is shown that the generalizations of the Weierstrass E-condition and the transversality conditions presented in the second paper remain essentially valid for an original control. The generalization is in the sense that the class of controls may, but need not, consist of relaxed controls.

INTRODUCTION

Paper No. 4

The fourth paper, written by S. K. Lakhanpal of Vanderbilt University, presents an application of Hamilton-Jacobi theory to a planar thrusting trajectory in a central force field. The paper presents the background theory needed to formulate and solve the "base" problem (thrust is equal to zero) and applies it in the problem using two different methods. The complete integral is obtained by Lagrange's linear equation in the first application and by Jacobi's method in the second application. In both applications Hamilton's equations are presented in the transformed variables.

Paper No. 5

The fifth paper, written by P. M. Fitzpatrick and J. E. Cochran of Auburn University, covers the use of Liouville's theorem for deriving a generating function for transformations of a Hamiltonian system. Methods are developed for making transformations which make use of the known constants of integration by putting the variables and constants in the form so that Liouville's theorem may be applied. The methods are then applied to two examples. The examples are the orbit in the central force field and free motion of a triaxial rigid body.

Paper No. 6

The sixth paper, written by C. F. Price of Massachusetts Institute of Technology, presents an offset vector iteration method for solving two-point, boundary-value problems along with a modification. The method depends on an "approximate solution". It has the distinct advantage of moving toward the desired solution with each pass through the ordinary differential equations of motion and, therefore, if the "approximate solution" gives a solution sufficiently near the desired end conditions, it may converge on the end conditions with far fewer passes through the ordinary differential equations than higher order methods. It is pointed out that the information generated may be stored for use by higher order iteration procedures, should this be desirable.

INTRODUCTION

Paper No. 7

The seventh paper, written by T. N. Edelbaum of Analytical Mechanics Associates, Inc., presents a linearized guidance procedure for a space trajectory. The space trajectory is a minimum-fuel trajectory and the thrusting is impulsive. The guidance corrections are impulsive and are designed to be valid in the neighborhood of the nominal trajectory. This paper covers three different problems; (1) the time-open rendezvous case, (2) the time-open orbit transfer, and (3) the time-open orbit transfer where one or more finite impulses are tangent to the velocity vector.

Paper No. 8

The eighth paper written by J. P. Vinti of Massachusetts Institute of Technology presents a set of equations for computing orbits in closed form using the spheroidal method (Vinti potential) of calculation. The equations are good in the general case and in particular they are good for polar and near polar orbits. The paper develops the changes in the known equations so that near the polar orbits division by differences of near equal quantities (near zero) is avoided. Thus, the numerical accuracy is enhanced. The procedure transforms the equations so that an explicit parameter for the right ascension does not appear. This is the troublesome variable.

Paper No. 9

The ninth paper, written by A. A. Kamel of Stanford University, presents procedures for developing expansion formulae in canonical transformations depending on a small parameter where the implementation of such perturbation theory is put in a form for speedy computerized symbolic manipulation. "Deprits' equations" are developed using a linear operator called Lie derivative generated by W, where the generating function W has a special form. Recursive relationships of the transformed variables and Hamiltonian are then developed. These recursive relationships are then modified by the introduction of intermediate functions to increase the speed of computerized symbolic manipulations.

INTRODUCTION

Paper No. 10

The tenth paper, written by P. Sconzo and D. Valenzuela of International Business Machines' Cambridge Advanced Space Systems Department, presents a formal solution of the n-body problem. This solution is a Taylor series in time for each of the 3n variables with coefficients generated recursively from the 6n initial conditions. It is obtained by a careful selection of intermediate variables and by the use of PL/1 FORMAC.

Paper No. 11

The eleventh paper, written by R. Dasenbrock of Stanford University, is on the long period behavior of a close lunar orbiter. A reference frame is chosen which is rotating with the moon with the x-axis in the equatorial plane determined by Cassini's law and the z-axis along the axis of rotation of the moon. The Hamiltonian is written in this rotating system in mixed Keplerian and Delaunay variables. The parts of the Hamiltonian are then ordered and integration of the equations is obtained with the short period terms averaged out by a series of canonical transformations. It is pointed out that there are 11 critical inclinations. Near these inclinations the von Zeipel method, which was used, fails. The case of the polar orbit is discussed separately. Phase plane contours of H "and h" with constant F** are presented and discussed. This work is a continuation of earlier work done by J. Vagners documented in NASA-ERC PM-67-21, pp. 213-228.

Paper No. 12

The twelth paper, written by W. T. Kyner of the University of Southern California, contains an exposition of the theory of non-linear resonance followed by application to the J_{22} perturbations on the orbit of a 24-hour synchronous satellite. The expository portion is based on lectures delivered at the 1968 Summer Institute of Dynamical Astronomy. In the application, it is shown that the longitude on a synchronous satellite satisfies a pendulum equation on the average. The validity of the pendulum model is restricted to time intervals of the order of $1/\sqrt{J_{22}}$.

INTRODUCTION

Two internal publications authored or co-authored by members of the sponsoring laboratory and in the subject technical fields have appeared since the last compilation. These are listed below with their summaries.

Miner, William E.: The Equations of Motion for Optimized Propelled Flight Expressed in Delaunay and Poincare Variables and Modifications of These Variables. NASA TN D-4478, May 1968.

SUMMARY

This document presents methods for developing the ordinary differential equations (o.d.e.) of motion in canonical form equivalent to the forms of Delaunay and Poincare. It also presents modifications to these forms so that three variables, which are constants of motion, result while the forms remain canonical.

The equations of motion are for a vehicle propelled by constant thrust magnitude with a constant mass flow rate. The vehicle is moving in a central force field. The trajectories are optimum in the sense of classical calculus of variations in a neighborhood definable by the boundary conditions of the specific problem. Specific problems are not discussed in this document.

The value of the document lies in two major areas:

1. The possible economics in numerical calculations which may result from using these ordinary differential equations, and

2. The application of the general perturbation theory of classical celestial mechanics to approximate solutions of these ordinary differential equations.

This document has been written to record the results of the investigation and was not meant to be a tutorial treatment of the subject. For such treatment, the references listed below are recommended by the author:

1. Bliss, G. A.: Lectures on the Calculus of Variations. University of Chicago Press, Chicago, Ill., 1961.
2. Goldstein, H.: Classical Mechanics. Addison-Wesley Publishing Co., Inc., Cambridge, Mass., March 1956.

3. Ford, L. R.: Differential Equations. McGraw-Hill Book Co., Inc., N. Y., 1933.
4. Smart, W. M.: Celestial Mechanics. Longmans, Green, and Co., Ltd., London, 1953.

****

Hoelker, R. F., and Winston, B. P.: A Comparison of a Class of Earth-Moon Orbits with a Class of Rotating Kepler Orbits. NASA TN D-4903, November 1968.

SUMMARY

In two concurrent series of graphs, a class of orbits in the Earth-Moon (E-M) field and a class of Kepler orbits in rotating coordinates are depicted and compared.

A general discussion of characteristics of rotating Kepler orbits is included.

The model used for the E-M orbits is that of the restricted problem of three bodies. The orbits of the class depicted originate at the E-M line, half of the E-M distance beyond the moon with velocity orthogonal to the E-M line within the E-M plane.

A GENERALIZED MULTISTAGE PROBLEM OF BOLZA IN THE CALCULUS OF VARIATIONS, I

By J. L. Linnstaedter
Associate Professor of Mathematics
Arkansas State University
State University, Arkansas

A GENERALIZED MULTISTAGE PROBLEM OF BOLZA IN THE CALCULUS OF VARIATIONS, I*

By J. L. Linnstaedter
Associate Professor of Mathematics
Arkansas State University
State University, Arkansas

SUMMARY

The problem is to find in a class of admissible arcs, satisfying certain multistage differential equations of constraint and end and intermediate point constraints, one which minimizes a Bolza type expression. The differential constraints may be different and different in number on the separate stages. Admissible arcs are continuous and piece-wise smooth in each stage but may be actually discontinuous at stage boundaries. The number of stages is bounded but otherwise not predetermined, since any stage will be allowed to degenerate to a null status. This is a generalization of the Denbow multistage extension of the Problem of Bolza. Appropriate imbedding theorems, a multiplier rule, and analogues of the Weierstrass and Clebsch conditions are obtained.

The theory of the second variation, the accessory minimum problem, and conjugate point conditions have been developed and will be presented in a subsequent paper.

INTRODUCTION

This study was motivated by the multistage character of many space-flight optimization problems. The problem treated is a generalization of the Denbow multistage extension of the Bolza problem [reference 3].

*This work was largely done at Vanderbilt University on NASA Research Grant NGR 43-002-015. The author wishes to thank Dr. M. G. Boyce for many helpful discussions during the performance of this research.

MULTISTAGE PROBLEM OF BOLZA

It is a generalization in the sense that the differential constraints may be different and different in number on the various stages, stages are allowed to degenerate, and discontinuities at staging points are permitted. The problem is approached directly as a multistage problem using extensions of the methods used on the Bolza problem [1]. This approach avoids the transformation to a Bolza problem used by Denbow [3].

Certain multistage control problems can be included in this problem by using techniques of Hestenese [4] and Valentine [6] as has been shown for a simpler case by Boyce and Linnstaedter [2]. The applicability of multistage variational problems is best illustrated in a recent paper by Miner and Andrus [5].

Three imbedding theorems, a multiplier rule, and analogues of the Weierstrass and Clebsch conditions are given. The first two imbedding theorems ignore the end and intermediate conditions and consider comparison arcs satisfying only the differential constraints. The necessary conditions given reduce to those for the Bolza problem whenever the problem degenerates to one stage.

FORMULATION OF THE PROBLEM

The problem is to find in a class of admissible arcs

$$y_i(x);\ x_o \le x_1 \le \cdots \le x_p;\ x \in [x_o, x_p];\ i = 1, 2, \ldots, n;$$

satisfying differential equations of constraint

$$\varphi_\beta^a(x, y, y') = 0;\ \beta = 1, 2, \ldots, m_a < n;\ a = 1, 2, \ldots, p;$$

$$x \in [x_{a-1}, x_a];$$

and end and intermediate point conditions

$$J_\mu [x_o, x_1, \ldots, x_p, y(x_o), y(x_1^-), y(x_1^+), \ldots, y(x_{p-1}^+), y(x_p)] = 0;$$

$$\mu = 1, 2, \ldots, q \le (2n + 1)p + 1;$$

one which minimizes a sum of the form

$$J = g[x_o,\ldots,x_p,y(x_o),y(x_1^-),y(x_1^+),\ldots y(x_p)] + \sum_{a=1}^{p}\int_{x_{a-1}}^{x_a} f^a(x,y,y')dx.$$

In the above statement and hereafter, y denotes the set $(y_1,\ldots,y_n)$, and primes indicate differentiation with respect to x. Require $y_i(x)$ to be continuous for $x \in [x_o, x_p] - (x_1, x_2, \ldots, x_{p-1})$ and $y_i'(x)$ to be piecewise continuous for $x \in [x_o, x_p]$, where i = 1, 2, ..., n. The finite non-decreasing set of points $(x_o, x_1, \ldots, x_p)$ will be called a set of partition or staging points. The $x_o, x_1, \ldots, x_p$ are not fixed but are to be determined by the minimization requirement. The left and right limits of y_i and y_i' at points of discontinuity are assumed to be defined and finite. Variables occuring as subscripts denote partial derivatives and repeated indices in a product indicate summation. Let R be an open connected set of 2n + 1 dimensional (x, y, y') space with φ^a_β, f^a having continuous third order partial derivatives in R. Furthermore, let the matrix $\left\|\varphi^a_{\beta y_i'}\right\|$ have rank m_a in R. Let S be an open set of 2np + p + 1 dimensional

$$(x_o, x_1,\ldots, x_p, y(x_o), y(x_1^-), y(x_1^+),\ldots,y(x_{p-1}^-),y(x_{p-1}^+), y(x_p))$$

space, with J_μ, g having continuous third order partial derivatives in S. Moreover, require the matrix

$$\left\|J_{\mu x_o}\ J_{\mu x_1}\ \cdots\ J_{\mu x_p}\ J_{\mu y(x_o)}\ J_{\mu y(x_1^-)}\ J_{\mu y(x_1^+)}\ \cdots\ J_{\mu y(x_p)}\right\|$$

to have rank q in S.

A set (x, y, y') is an admissible set if it is contained in R. An admissible subarc C^a is a set of functions $(y_1, y_2, \ldots, y_n)$, $x \in [x_{a-1}, x_a]$ with (x, y, y') an admissible set and such that y_i is continuous and y_i' is piecewise continuous on $[x_{a-1}, x_a]$ for each

i, $i = 1, 2, \ldots, n$. An admissible arc C is a partition set $(x_o, x_1, \ldots, x_p)$ together with a set of admissible subarcs C^a, $a = 1, 2, \ldots, p$, such that the set $(x_o, \ldots, x_p, y(x_o), y(x_1^-), y(x_1^+), \ldots, y(x_p)) \in S$. On each admissible arc, φ_β^a, f^a, J, g, J_μ are assumed to be defined.

ADMISSIBLE FAMILIES AND VARIATIONS

Suppose there exists an admissible arc E satisfying $\varphi_\beta^a = J_\mu = 0$. If there are no other arcs satisfying $\varphi_\beta^a = J_\mu = 0$ with which to compare E then the problem is trivial. In order to establish that the problem is not trivial, we will give conditions that an admissible arc E can be imbedded in a family of comparison arcs. This will be the content of Theorems 1, 2, and 4. Theorem 4 gives conditions that guarantee other arcs in a neighborhood of E that satisfy $\varphi_\beta^a = 0$ and $J_\mu = 0$ while Theorems 1 and 2 guarantee other arcs near E satisfying only $\varphi_\beta^a = 0$. First, we need the following definitions, the first two being essentially the same as are given in Bliss and the third one is a multistage extension of the definition of admissible family given in Bliss [1, 194-195]. A one-parameter family of arcs $y_i(x, b)$; $x' \leq x \leq x''$, $|b| \leq \varepsilon$; is an elementary family if and only if $y_i'(x, b)$ exist and $y_i(x, b)$ have continuous first derivatives with respect to b in a neighborhood of points (x, b) containing $x' \leq x \leq x''$, $|b| \leq \varepsilon$. Two elementary families are said to be adjacent if and only if they are defined on adjacent intervals and are continuous across the common end point. These definitions hold between partition points but not necessarily across partition points. A family of arcs will be called an admissible family if and only if $y_i(x,b)$ exist for $x_o(b) \leq x \leq x_p(b)$, $|b| < \varepsilon$; $x_o(b)$, $x_1(b)$, $\ldots$, $x_p(b)$ have continuous first derivatives with respect to b in the region $|b| < \varepsilon$;

for each a there is a finite sequence of intervals $[x', x'']$, $[x'', x''']$, ..., $[x^{(k-1)}, x^{(k)}]$ for k depending on a such that $x' < x_{a-1}(b) < x''$ and $x^{(k-1)} < x_a(b) < x^{(k)}$; $y_i(x, b)$ for $x \in [x_{a-1}, x_a]$ is a part of a finite sequence of adjacent elementary families belonging to the sequence of intervals.

The notation to be used for differentials of an admissible family is as follows:

$$dx_o = x_{ob}db,\ dx_1 = x_{1b}db,\ \ldots,\ dx_p = x_{pb}db;$$

$$dy_i = y_i'dx + \delta y_i \text{ where } \delta y_i = y_{ib}db \text{ and } y_i' = y_{ix}.$$

The set of variations of the family along the arc E is the set $\xi_o, \xi_1, \ldots, \xi_p, \eta_i(x)$ defined by

$$dx_o = x_{ob}(0)\ db = \xi_o db,\ \ldots,\ dx_p = x_{pb}(0)\ db = \xi_p db;$$

$$\delta y_i = y_{ib}(x, 0)\ db = \eta_i(x)\ db.$$

The $\xi_o, \xi_1, \ldots, \xi_p$ are constants and the $\eta_i(x)$ are continuous and have piecewise continuous derivatives between partition points of E. Every set $\xi_o, \xi_1, \ldots, \xi_p, \eta_i(x)$ with these properties is called a set of admissible variations along E.

If we require the arcs of an admissible family to satisfy $\varphi_\beta^a = 0$, then the variations $\eta_i(x)$ along E contained in the family for $b = 0$ satisfy

$$\Phi_\beta^a(x, \eta, \eta') = \varphi^a_{\beta y_i}\eta_i + \varphi^a_{\beta y_i'}\eta'_i = 0,$$

where the arguments of $\varphi^a_{\beta y_i}$ and $\varphi^a_{\beta y_i'}$ are $(x, y(x, 0), y'(x, 0))$ belonging to E. The equations $\Phi_\beta^a = 0$ are called the equations of variation along E. In these equations repeated subscripts indicate

summation. If E is specified then the coefficients of η_i and η_i' are fixed and independent of any family.

The equations of variation on E of the end and intermediate point conditions will be given by

$$\hat{J}_\mu = J_{\mu x_o}\, \xi_o + \cdots + J_{\mu x_p}\, \xi_p + J_{\mu y_i(x_o)}\, \eta_i(x_o) + J_{\mu y_i(x_1^-)}\, \eta_i(x_1^-) +$$

$$J_{\mu y_i(x_1^+)}\, \eta_i(x_1^+) + \cdots + J_{\mu y_i(x_p)}\, \eta_i(x_p)$$

where the arguments of the coefficients of the variations are the end and intermediate values of E.

IMBEDDING THEOREMS

We can now state the first imbedding theorem. The proof of this theorem is a specialization of the proof of Theorem 2 and for this reason it is omitted.

Theorem 1. If an admissible arc E satisfies the equation $\varphi_\beta^a = 0$, and if $\xi_o, \xi_1, \cdots, \xi_p, \eta_i(x)$ is a set of admissible variations satisfying the equations of variation $\overset{1}{\phi}{}_\beta^a = 0$ on E, then there is a one-parameter admissible family $y_i(x, b)$ of arcs containing E, for the parameter value $b = 0$, satisfying the equations $\varphi_\beta^a = 0$, and having the set $\xi_o, \xi_1, \cdots, \xi_p, \eta_i(x)$ as the variations of the family along E.

The extension of this theorem to an s-parameter family is the content of the following theorem.

Theorem 2. If an admissible arc E satisfies the equations $\varphi_\beta^a = 0$ and if $\xi_{o\sigma}, \xi_{1\sigma}, \cdots, \xi_{p\sigma}, \eta_{io}(x)$, $(\sigma = 1, 2, \ldots, s)$ are s sets of admissible variations satisfying the equations of variation $\overset{1}{\phi}{}_\beta^a = 0$

along E, then there is an admissible s-parameter family $y_i(x,b_1,b_2,\ldots b_s)$ of arcs containing E for the parameter values $b_\sigma = 0$ ($\sigma = 1, 2, \ldots, s$), satisfying the equations $\varphi^a_\beta = 0$, and having for each $\sigma = 1, \ldots, s$ the set $\xi_{0\sigma}, \xi_{1\sigma}, \ldots, \xi_{p\sigma}, \eta_{i\sigma}(x)$ as the variations of the family along E with respect to the parameter b_σ.

Proof. Let E be an admissible arc satisfying $\varphi^a_\beta = 0$, and let $\xi_{0\sigma}, \xi_{1\sigma}, \ldots, \xi_{p\sigma}, \eta_{i\sigma}(x)$, ($\sigma = 1, 2, \ldots, s$) be s sets of admissible variations satisfying the equations of variations $\overset{1}{\varphi}{}^a_\beta = 0$ along E. Consider any arbitrary non-degenerate stage a with associated partition interval $[x_{a-1}, x_a]$. Extend the system of equations $\varphi^a_\beta(x,y,y') = 0$ by introducing new equations $z_\gamma = \varphi^a_\gamma$, ($\gamma = m_a + 1, \ldots, n$), where the functions $\varphi^a_\gamma(x,y,y')$ are chosen so as to have continuous partial derivatives of at least third order in a neighborhood of the values (x,y,y') belonging to E^a and such that $\left|\varphi^a_{\gamma y_i'}\right| \neq 0$ along E^a, $\gamma = 1, 2, \ldots, m_a, m_a+1, \ldots, n$. The z_γ are new variables, and E^a is the subarc of E associated with the a stage. The equations $\varphi^a_\beta = 0$, $\varphi^a_\gamma = z_\gamma$ determine functions $z_\gamma(x)$ belonging to E^a when $y_i(x)$ defining E^a are substituted in these equations. The $z_\gamma(x)$ are continuous except possibly at corners of E^a. The equations of variations are

$$\overset{1}{\varphi}{}^a_\beta = 0, \qquad \overset{1}{\varphi}{}^a_\gamma = \mathcal{J}_\gamma$$

where the functions of $\mathcal{J}_\gamma$ are variations of z_γ associated with the subarc E^a and the variations η_i. The $\mathcal{J}_\gamma$ are dependent on η_i and η_i', so $\mathcal{J}_{i\sigma}$ corresponds to $\eta_{i\sigma}$, $\sigma = 1, 2, \ldots, s$. Furthermore, for each σ, $\mathcal{J}_{i\sigma}(x)$ is continuous except possibly at corners of E^a or discontinuities of $\eta'_{i\sigma}(x)$.

The extended system of differential equations has solutions $y_i' = M_i(x, y, z)$ with M_i having continuous partial derivatives of at least third order in a neighborhood of the values (x, y, z) belonging to E^a, since φ_β^a, φ_γ^a have continuous third order partials. Let x' be the first value of x following x_{a-1} defining a corner of E^a or a discontinuity of $\eta_{i\sigma}'(x)$, or let $x' = x_a$ if there are no corners on E^a or discontinuities in $\eta_{i\sigma}'(x)$. The functions z_γ, $\mathcal{f}_{\gamma\sigma}$ as defined on $[x_{a-1}, x']$ can be extended so that they are continuous on a slightly larger interval. The right members of the equations

$$y_i' = M_i(x, y_i, z_\gamma(x) + b_\sigma \mathcal{f}_{\gamma\sigma}(x))$$

are continuous in $x, y_1, \ldots, y_n, b_1, \ldots, b_s$ and have continuous third partial derivatives with respect to the variables $y_1, \ldots, y_n, b_1, \ldots, b_s$ in a neighborhood of the values $x, y_1, \ldots, y_n, b_1 = 0, \ldots, b_s = 0$ belonging to E_1^a where E_1^a is the subarc of E^a associated with $[x_{a-1}, x']$. Solutions

$$y_i = Y_i(x, \bar{x}, \bar{y}, b_1, \ldots, b_s)$$

exist for initial point $(\bar{x}, \bar{y}_1, \ldots, \bar{y}_n)$ with Y_i, Y_i' continuous and having continuous partial derivatives of at least third order with respect to the arguments y_i, b_σ in a neighborhood of the sets $(x, \bar{x}, \bar{y}_i, b_\sigma)$ belonging to E^a.

The functions

$$y_i = Y_i[x, x_{a-1}, y_i(x_{a-1}^+) + b_\sigma \eta_{i\sigma}(x_{a-1}^+), b_\sigma] = y_i(x, b_\sigma)$$

define an elementary family satisfying the equations $\varphi_\beta^a = 0$ on an interval including $[x_{a-1}, x']$.

The functions $y_i(x, b_\sigma)$ have at x_{a-1} the initial values

$$y_i(x^+_{a-1}, b_\sigma) = Y_i[x_{a-1}, x_{a-1}, y_i(x^+_{a-1}) + b_\sigma \eta_{i\sigma}(x^+_{a-1}), b_\sigma]$$

$$= y_i(x^+_{a-1}) + b_\sigma \eta_{i\sigma}(x^+_{a-1}).$$

Furthermore, $y_{ib_\sigma}(x, 0)$ along E^a_1 have at $x = x_{a-1}$ the initial values $\eta_{i\sigma}(x^+_{a-1})$ and $y_i(x, b_\sigma)$ satisfies $\varphi^a_\beta = 0$, $\varphi^a_\gamma = z_\gamma(x) + b_\sigma \mathcal{J}_{\gamma\sigma}(x)$. Thus along E^a_1, $y_{ib_\sigma}(x, 0) = \eta_{i\sigma}(x)$ because of the uniqueness of solutions with initial values $\eta_{i\sigma}(x^+_{a-1})$.

This determines an s-parameter elementary family on the interval $[x_{a-1}, x']$ satisfying $\varphi^a_\beta = 0$ and having $y_{ib_\sigma}(x, 0) = \eta_{i\sigma}(x)$ along E^a_1. Let x'' be the next value of x following x' on $[x_{a-1}, x_a]$ defining a corner of E^a or a discontinuity of $\eta'_{i\sigma}(x)$, or x_a if E^a has no other corners and $\eta'_{i\sigma}(x)$ have no other discontinuities. Repeating the preceding arguments produces an s-parameter elementary family on $[x', x'']$ which is adjacent to the elementary family on $[x_{a-1}, x']$ and satisfying $\varphi^a_\beta = 0$ with $y_{ib_\sigma}(x, 0) = \eta_{i\sigma}(x)$ on E^a_2 (subarc for $[x', x'']$). Continuing this process for a finite number of times gives a finite sequence of adjacent s-parameter elementary families which together give an s-parameter family of arcs in R satisfying the properties of the theorem for the a stage with

$$x_{a-1}(b_\sigma) = x_{a-1} + b_\sigma \mathcal{J}_{a-1\sigma}, \quad x_a(b_\sigma) = x_a + b_\sigma \mathcal{J}_{a\sigma}.$$

By identifying the parameters of each stage with those of adjacent stages, an s-parameter family satisfying the requirements of the theorem is obtained.

THE FIRST VARIATION

Let an admissible arc E be imbedded in a one-parameter family

$y_i(x, b)$ with E determined by $y_i(x, 0)$. Evaluate J along the family so that J is a function of the parameter b as follows:

$$J(b) = g[x_o(b), x_1(b), \ldots, x_p(b), y_i(x_o(b), b), y_i(x_1^-(b), b), \ldots, y_i(x_p(b), b)]$$

$$+ \sum_{a=1}^{p} \int_{x_{a-1}(b)}^{x_a(b)} f^a(x, y(x, b), y'(x, b))dx.$$

Taking the differential of J, we have

$$dJ = dg + \sum_{a=1}^{p} f^a dx \Big|_{x_{a-1}}^{x_a} + \sum_{a=1}^{p} \int_{x_{a-1}}^{x_a} (f^a_{y_i} \delta y_i + f^a_{y_i'} \delta y_i')dx.$$

The <u>first variation</u> of J along E is $\hat{J}$ where $dJ = \hat{J}db$ with dJ evaluated along E. Explicitly,

$$\hat{J}(\xi, \eta) = \hat{g} + \sum_{a=1}^{p} f^a \xi \Big|_{a-1}^{a} + \sum_{a=1}^{p} \int_{x_{a-1}}^{x_a} (f^a_{y_i} \eta_i + f^a_{y_i'} \eta_i')dx$$

where the arguments of f^a, $f^a_{y_i}$, $f^a_{y_i'}$ are those determining E and

$$\hat{g} = g_{x_o} \xi_o + g_{x_1} \xi_1 + \cdots + g_{x_p} \xi_p + g_{y_i(x_o)} [y_i'(x_o)\xi_o + \eta_i(x_o)]$$

$$+ g_{y_i(x_1^-)} [y_i'(x_1^-) \xi_1 + \eta_i(x_1^-)] + g_{y_i(x_1^+)} [y_i'(x_1^+) \xi_1 + \eta_i(x_1^+)]$$

$$+ \cdots + g_{y_i(x_p)} [y_i'(x_p) \xi_p + \eta_i(x_p)].$$

Define F^a as follows:

$$F^a(x, y, y', \lambda) = \lambda_o f^a + \lambda_\alpha \varphi^a_\alpha$$

with λ_o a constant, λ_α a function of x for each α, and φ^a_α are functions φ^a_β, φ^a_γ described earlier. Since

$$\lambda_\beta \Phi_\beta + \lambda_\gamma (\Phi_\gamma - \mathcal{J}_\gamma) = 0,$$

it can be added to the integrand of $\lambda_o \hat{J}$ without changing the integrand's value. Thus

$$\lambda_o \hat{J}(\xi, \eta) = \lambda_o \hat{g} + \sum_{a=1}^{p} \lambda_o f^a f \Big|_{a-1}^{a} + \sum_{a=1}^{p} \int_{x_{a-1}}^{x_a} [F^a_{y_i} \eta_i + F^a_{y_i'} \eta_i' - \lambda_\gamma f_\gamma] dx.$$

To prove the multiplier rule, we need the following lemma.

Lemma 1. Let $x \in [x_{a-1}, x_a]$ and $x_{a-1} \neq x_a$. If λ_o, c_i $(i = 1, 2, \ldots, n)$ are arbitrarily selected constants then there are multipliers $\lambda_\alpha(x)$, determined uniquely by

$$F^a_{y_i'} = \int_{x_{a-1}}^{x} F^a_{y_i} \, dx + c_i,$$

which are continuous except possibly at corners of E^a.

Proof. Following procedures of Bliss for the Bolza problem, define

$$v_i = F^a_{y_i'} = \lambda_o f^a_{y_i'} + \lambda_\alpha \varphi^a_{\alpha y_i'}.$$

Now consider

$$dv_i/dx = F^a_{y_i} = \lambda_o f^a_{y_i} + \lambda_\alpha \varphi^a_{\alpha y_i}$$

and

$$v_i(x_{a-1}) = c_i.$$

Further notice that the first system of equations can be solved for $\lambda_\alpha(x)$ in terms of λ_o and $v_i(x)$. Substituting these in the system of differential equations gives

$$dv_i/dx = A_{i\alpha} v_\alpha + \lambda_o B_i.$$

The coefficients $A_{i\alpha}$, B_i are continuous functions of x between corners of E^a. The existence of continuous (between corners) solutions $v_i(x)$ of this system with initial conditions $v_i(x_{a-1}) = c_i$ is equivalent to

finding continuous functions $\lambda_i(x)$ (between corners) for the integral equations of the lemma. The system

$$dv_i/dx = A_{i\alpha}\, v_\alpha + \lambda_o B_i, \quad v_i(x_{a-1}) = c_i$$

is a linear first order system of differential equations. Let x' be the first corner of E^a following x_{a-1}. The initial condition then for the solution between corners defined by x' and x'' is simply $v_i(x')$. Continuing this it is clear that the system has continuous solutions $v_i(x)$ (between corners) and hence there are continuous solutions $\lambda_\alpha(x)$ (between corners) determined by the integral equations of the lemma. The uniqueness of solutions to the system of differential equations through a fixed point guarantees the uniqueness of the multipliers $\lambda_\alpha(x)$. The set λ_o, $v_i(x)$, and consequently the set λ_o, $\lambda_\alpha(x)$, do not vanish simultaneously at a point unless they are all identically zero.

We get a reduced form for the first variation by using integration by parts and the integral equations of this lemma. Hence, for all admissible variations ξ_o, ξ_1, $\cdots$, ξ_p, $\eta_i(x)$ satisfying the equations $\bar{\Phi}^a_\beta = 0$, we have

$$ {}_oJ(\xi, \eta) = \lambda_o \sum_{a=1}^{p} f^a \xi \Big|_{a-1}^{a} + \lambda_o \hat{g} + \sum_{a=1}^{p} F^a_{y_i}\, \eta_i \Big|_{a-1}^{a} $$

$$ - \sum_{a=1}^{p} \int_{x_{a-1}}^{x_a} \lambda_\gamma \xi_\gamma \, dx. $$

MULTIPLIER RULE

We now proceed to state and prove a multiplier rule, or first necessary condition.

Theorem 3. An admissible arc E, defined on an interval $[x_o, x_p]$ is said to satisfy the multiplier rule if there exist constants λ_o, e_μ, not all zero, and a set of functions

$$F^a(x, y, y', \lambda) = \lambda_o f^a + \lambda_\beta(x)\, \varphi^a_\beta, \quad a = 1, \ldots, p,$$

with multipliers $\lambda_\beta(x)$ continuous on $[x_o, x_p]$ except possibly at partition points and corners of E, such that the equations

$$F^a_{y'_i} = \int_{x_{a-1}}^{x} F^a_{y_i}\, dx + c^a_i, \quad \varphi^a_\beta = 0, \quad x \in [x_{a-1}, x_a],$$

are satisfied on E and the equations

$$\sum_{a=1}^{p} \left\{F^a - F^a_{y'_i}\, y'_i\right\} dx \Big|_{a-1}^{a} + \sum_{a=1}^{p} F^a_{y'_i}\, dy_i \Big|_{a-1}^{a}$$

$$+ \lambda_o\, dg + e_\mu\, d J_\mu = 0, \quad J_\mu = 0$$

hold for end and intermediate points of E for every choice of differentials $dx_o, dx_1, \ldots, dx_p, dy_i(x_o), dy_i(x_1^-), dy_i(x_1^+), \ldots, dy_i(x_p)$. For an arc E satisfying the multiplier rule the multipliers λ_o, $\lambda_\beta(x)$ do not vanish simultaneously at any point of $[x_o, x_p]$, and right and left limits are defined at partition points and corners of E. Every minimizing arc E must satisfy the multiplier rule.

Proof. Let E be a minimizing arc for this problem. Let $\eta_{o\tau}$, $\eta_{1\tau}$, $\ldots$, $\eta_{p\tau}$, $\eta_{i\tau}(x)$ be q + 1 sets of admissible variations all of which satisfy the equations of variations $\Phi^a_\beta = 0$ along E, $\tau = 0, 1, \ldots, q$. By Theorem 2, there is a (q + 1) - parameter admissible family of arcs

$$y_i\ (x, b_o, b_1, \ldots, b_q)$$

containing E for $b_\tau = 0$, satisfying $\varphi^a_\beta = 0$, and having $\eta_{o\tau}$, $\eta_{1\tau}$, $\ldots$,

$\xi_{p\tau}$, $\eta_{i\tau}(x)$ as its variations with respect to b_τ along E. The functions J, J_μ become functions of the parameters when the functions defining the (q + 1) - admissible family are substituted in them. The equations

$$J(b_o, b_1, \ldots, b_q) = J(0, \ldots, 0) + u,$$

$$J_\mu(b_o, b_1, \ldots, b_q) = 0$$

have the solution $b_o = b_1 = \ldots = b_q = u = 0$ corresponding to the minimizing arc E. We wish to show the determinant

$$\begin{vmatrix} \frac{\partial J}{\partial b_o} & \cdots & \frac{\partial J}{\partial b_q} \\ \\ \frac{\partial J_\mu}{\partial b_o} & \cdots & \frac{\partial J_\mu}{\partial b_q} \end{vmatrix}$$

is zero for $b_o = b_1 = \ldots = b_q = u = 0$. Suppose the contrary, then from existence theorems for implicit functions the above equations have unique solutions b_τ (u) continuous near u = 0 and having initial values $b_\tau(0) = 0$. But for negative values of u, the value of J on E is larger than the value of J along some admissible arc corresponding to $b_\tau(u)$ (for negative u). This contradicts the fact that J takes on its minimum value along E. Hence the determinant is zero for all choices of the variations and takes the form

$$\begin{vmatrix} \hat{J}(\xi_{co}, \eta_{io}) & \cdots & \hat{J}(\xi_{cq}, \eta_{iq}) \\ \\ \hat{J}_\mu(\xi_{co}, \eta_{io}) & \cdots & \hat{J}_\mu(\xi_{cq}, \eta_{iq}) \end{vmatrix}$$

where $c = 0, 1, \ldots, p$ and $i = 1, \ldots, n$ and $\mu = 1, \ldots, q$.

We notice that this determinant is $q + 1$ by $q + 1$. Let $t < q + 1$ be the maximum rank attainable for this determinant for sets of admissible variations $\xi_{0\tau}, \xi_{1\tau}, \ldots, \xi_{p\tau}, \eta_{i\tau}(x)$ satisfying $\Phi^a_\beta = 0$ on E. Furthermore, let this set of admissible variations be a set for which this rank is attained. Thus, there exist constants λ_0, e_μ (not all zero) satisfying the following system of equations:

$$\lambda_0 \hat{J}(\xi_{c0}, \eta_{i0}) + e_\mu \hat{J}_\mu(\xi_{c0}, \eta_{i0}) = 0,$$
$$\cdots\cdots\cdots$$
$$\lambda_0 \hat{J}(\xi_{cq}, \eta_{iq}) + e_\mu \hat{J}_\mu(\xi_{cq}, \eta_{iq}) = 0.$$

Now, with these constants the equation

$$\lambda_0 \hat{J}(\xi_c, \eta_i) + e_\mu \hat{J}_\mu(\xi_c, \eta_i) = 0$$

must be satisfied for every set of admissible variations $\xi_0, \xi_1, \ldots, \xi_p, \eta_i(x)$ satisfying the equations of variation $\Phi^a_\beta = 0$ along E. If this were not the case then there would exist a set of admissible variations $\xi^*_0, \ldots, \xi^*_p, \eta^*_i(x)$ such that

$$\lambda_0 \hat{J}(\xi^*_c, \eta^*_i) + e_\mu \hat{J}_\mu(\xi^*_c, \eta^*_i) \neq 0.$$

We notice then that the $q + 1$ by $q + 2$ matrix

$$\left\| \begin{matrix} \hat{J}(\xi_{c0}, \eta_{i0}) & \cdots & \hat{J}(\xi_{cq}, \eta_{iq}) & \hat{J}(\xi^*_c, \eta^*_i) \\ \hat{J}_\mu(\xi_{c0}, \eta_{i0}) & \cdots & \hat{J}_\mu(\xi_{cq}, \eta_{iq}) & \hat{J}_\mu(\xi^*_c, \eta^*_i) \end{matrix} \right\|$$

has rank $t + 1$ since otherwise

$$\lambda_0 \hat{J}(\xi^*, \eta^*) + e_\mu \hat{J}_\mu(\xi^*, \eta^*) = 0.$$

But this would contradict t being the maximum rank.

Substituting the simplified version $\lambda_o \hat{J}$ given just before the statement of the multiplier rule in $\lambda_o \hat{J} + e_\mu \hat{J}_\mu = 0$, we get

$$\lambda_o \sum_{a=1}^{p} f^a \mathcal{f} \Big|_{a-1}^{a} + \lambda_o \hat{g} + \sum_{a=1}^{p} F^a_{y_i}, \eta_i \Big|_{a-1}^{a}$$

$$+ e_\mu \hat{J}_\mu - \sum_{a=1}^{p} \int_{x_{a-1}}^{x_a} \lambda_\gamma \mathcal{f}_\gamma \, dx = 0.$$

The expressions not under the integrals are linear in $\mathcal{f}_o$, $\mathcal{f}_1$, ..., $\mathcal{f}_p$, $\eta_i(x_o)$, $\eta_i(x_1^-)$, $\eta_i(x_1^+)$, ..., $\eta_i(x_p)$. Consider the coefficients of $\eta_i(x_o)$, namely

$$\lambda_o g_{y_i(x_o)} - F^1_{y_i'}(x_o, y(x_o), y'(x_o), \lambda(x_o)) + e_\mu J_{\mu y_i(x_o)},$$

and recall Lemma 1. Since we can select $c_i^1 = F^1_{y_i'}(x_o, y(x_o), y'(x_o), \lambda(x_o))$ arbitrarily, we can make this coefficient vanish by simply setting $c_i^1 = -\lambda_o g_{y_i(x_o)} + e_\mu J_{\mu y_i(x_o)}$. Similar remarks can be made concerning the vanishing of the coefficients of $\eta_i(x_1^+)$, $\eta_i(x_2^+)$, $\eta_i(x_3^+)$, ..., $\eta_i(x_{p-1}^+)$. The remaining expressions must vanish for every arbitrarily selected set

$$\mathcal{f}_o, \mathcal{f}_1, \ldots, \mathcal{f}_p, \eta_i(x_1^-), \eta_i(x_2^-), \ldots, \eta_i(x_{p-1}^-), \eta_i(x_p), \mathcal{f}_\gamma(x).$$

By choosing $\mathcal{f}_o = \ldots = \mathcal{f}_p = \eta_i(x_1^-) = \ldots = \eta_i(x_p) = 0$ we have

$$\sum_{a=1}^{p} \int_{x_{a-1}}^{x_a} \lambda_\gamma \mathcal{f}_\gamma \, dx = 0.$$

Now the λ_γ, $\gamma = m_a + 1, \ldots, n$ must vanish identically since the $\mathcal{f}_\gamma$ can be arbitrarily selected. Similar choosing will show that the coefficients of $\mathcal{f}_o, \mathcal{f}_1, \ldots, \mathcal{f}_p, \eta_i(x_1^-), \eta_i(x_2^-), \ldots, \eta_i(x_p)$

vanish. For $\varphi^a_\beta = 0$, $F^a = \lambda_o f^a$ and we can summarize by saying that the coefficients of $\xi_o, \ldots, \xi_p, \eta_i(x_o), \eta_i(x_1^-), \eta_i(x_1^+), \ldots, \eta_i(x_p)$ vanish in the following equation:

$$\sum_{a=1}^{p} F^a \xi \Big|_{a-1}^{a} + \lambda_o \hat{g} + \sum_{a=1}^{p} F^a_{y_i} \eta_i \Big|_{a-1}^{a} + e_\mu \hat{J}_\mu = 0.$$

Since $dx_o = \xi_o db$, $\ldots$, $dx_p = \xi_p db$, $dy_i(x_o) = y_i'(x_o)dx_o + \eta_i(x_o)db$, $dy_i(x_i^-) = y_i'(x_1^-)dx_1 + \eta_i(x_1^-)db$, $dy_i(x_1^+) = y_1'(x_1^+)dx_1 + \eta_i(x_1^+)db$, $\ldots$, $dy_i(x_p) = y_i'(x_p)dx_p + \eta_i(x_p)db$, this last equation can be transformed to the form given in the theorem. This together with Lemma 1 completes the proof of the multiplier rule.

COROLLARIES TO MULTIPLIER RULE

There are three important corollaries to this theorem.

Corollary 1. At each point between partition points of an admissible arc E satisfying

$$\varphi^a_\beta = 0, \quad F^a_{y_i'} = \int_{x_{a-1}}^{x} F^a_{y_i} \, dx + c^a_i, \qquad x \in [x_{a-1}, x_a],$$

the functions $F^a_{y_i'}$ have forward and backward derivatives, equal except at corner points and such that

$$dF^a_{y_i'}/dx = F^a_{y_i}.$$

Corollary 2. At each corner between partition points of an admissible arc E satisfying the equations in the hypothesis of Corollary 1, the functions $F^a_{y_i'}$ have defined left and right limits which are equal.

Corollary 3. On each sub-arc between partition points of an admissible arc E satisfying the equations in the hypothesis of Corollary 1, on which the functions $y_i(x)$ defining E have continuous derivatives and the determinant

$$R_a^* = \begin{vmatrix} F^a_{y_i' y_k'} & \varphi^a_{\beta', y_i'} \\ \varphi^a_{\beta y_k'} & 0 \end{vmatrix} \qquad \begin{matrix} (i, k = 1, \ldots, n) \\ (\beta, \beta' = 1, \ldots, m_a) \end{matrix}$$

is different from zero, the functions $y_i'(x)$, $\lambda_\beta(x)$ belonging to E have continuous derivatives of at least the first order with respect to x.

Corollaries 1 and 2 follow directly from the equations in the hypotheses.

Proof of Corollary 3. Let $\bar{x}$ be a value defining a point interior to some sub-arc of the a stage with the functions $y_i(x)$ defining E having continuous derivatives on this subarc and $R_a^* \neq 0$ at $\bar{x}$. The equations

$$F^a_{y_i'}(x, y(x), u(x), \mu(x)) = \int_{x_{a-1}}^{x} F^a_{y_i}(x, y(x), y'(x), \lambda(x)]dx + c_i^a,$$

$$\varphi^a_\beta(x, y(x), u(x)) = 0, \quad x \in [x_{a-1}, x_a],$$

have the solutions $u_i(x) = y_i'(x)$, $\mu_\beta(x) = \lambda_\beta(x)$ along E. Notice the R_a^* is the functional determinant of the left members of these equations with respect to u_i, μ_β and $R_a^* \neq 0$ at $(\bar{x}, y'(\bar{x}), \lambda(\bar{x})) = (\bar{x}, u(\bar{x}), \mu(\bar{x}))$. Theorems on implicit functions say that solutions $u_i = y_i'(x)$, $\mu_\beta = \lambda_\beta(x)$ will have continuous

derivatives with respect to x near $\bar{x}$ of as many orders as the functions in the equations, in this case φ^a_β, $F^a_{y'_i}$, have with respect to x, u, μ. This guarantees then that we have continuous derivatives of $y'_i(x)$, $\lambda_\beta(x)$ with respect to x of at least first order.

EXTREMALS, NORMAL AND ABNORMAL ARCS

A <u>stage extremal</u> for the <u>a</u> stage is an admissible subarc $y_i(x)$ without corners and with multipliers

$$\lambda_o = 1, \quad \lambda_\beta(x), \ \beta = 1, 2, \ldots, m_a, \ x \in [x_{a-1}, x_a],$$

for which $y'_i(x)$, $\lambda_\beta(x)$ have continuous derivatives on the interval $[x_{a-1}, x_a]$ and satisfy equations $\varphi^a_\beta = 0$ and $dF^a_{y'_i}/dx = F^a_{y_i}$. An <u>extremal</u> is an admissible arc which on each stage is a stage extremal. An <u>a stage extremal</u> is called <u>non-singular</u> provided $R^*_a \neq 0$ along it. An <u>extremal</u> is called <u>non-singular</u> if each of its stage extremals is non-singular.

Let M be the class of admissible arcs satisfying $\varphi^a_\beta = 0$, $J_\mu = 0$. An arc $E \in M$ is said to have <u>abnormality of order r</u> if it satisfies Theorem 3 (the multiplier rule) with r and only r linearly independent sets of multipliers of the form $\lambda_{o\rho} = 0$, $\lambda_{\beta\rho}(x)$, $\rho = 1, 2, \ldots, r$. If r = 0, the arc E is said to be <u>normal</u>. A set of multipliers λ_o, $\lambda_\beta(x)$ with $\lambda_o = 0$ will be an <u>abnormal set of multipliers</u>. If a normal arc E has a set of multipliers, then by dividing by a suitable constant these will have the form $\lambda_o = 1$, $\lambda_\beta(x)$. The set of multipliers with $\lambda_o = 1$ for a normal arc is unique, since if it had more than one they could be put in the form having $\lambda_o = 1$, and the

difference of two such sets of multipliers would be a set of multipliers with $\lambda_o = 0$ and hence abnormal.

A THIRD IMBEDDING THEOREM

We can now prove for multistage problems the theorem given by Bliss for one-stage problems [1, 214]. The proof again follows the pattern of Bliss' proof.

Theorem 4. If an arc $E \in M$ is normal, then there exists an admissible one-parameter family of arcs in M which includes E for parameter value $b = 0$ and which has in every neighborhood of E arcs of M not identical with E.

Proof. By Theorem 2, the normal arc E may be imbedded in an admissible $(q + 1)$ - parameter family of arcs $x_c(b_o, b_1, \ldots, b_q)$, $y_i(x, b_o, b_1, \ldots, b_q)$ satisfying only the differential equations $\varphi_\beta^a = 0$.

Consider the matrix

$$\left\| \begin{matrix} \hat{J}(\xi_{co}, \eta_{io}) & \cdots & \hat{J}(\xi_{cq}, \eta_{iq}) \\ \hat{J}_\mu(\xi_{co}, \eta_{io}) & \cdots & \hat{J}_\mu(\xi_{cq}, \eta_{iq}) \end{matrix} \right\|$$

and note that the maximum rank attainable for the last q rows must be q. For, if it were less than q, then there would be a set of constants $\lambda_o = 0$, e_μ (not all zero) satisfying the equation

$$\lambda_o \hat{J}(\xi_c, \eta_i) + e_\mu \hat{J}_\mu(\xi_c, \eta_i) = 0$$

for every set of admissible variations ξ_c, $\eta_i(x)$ and determining a set of multipliers $\lambda_o = 0$, $\lambda_\beta(x)$ (not vanishing simultaneously) for E. This contradicts E being normal. Now suppose that the variations

$\xi_{c\tau}$, $\eta_{i\tau}$ have been chosen so that the determinant of the first q columns of the last q rows of the above matrix is different from zero, and let these be the variations of the family $x_c\ (b_o, \ldots, b_q)$, $y_i(x, b_o, \ldots, b_q)$. Substitute into the functions J_μ, replace b_q by b, and consider the equations

$$J_\mu\ (b_o, b_1, \ldots, b_{q-1}, b) = 0.$$

These equations have the solution $b_o = b_1 = \ldots = b_{q-1} = b = 0$ at which the determinant

$$\left| \frac{\partial J_\mu}{\partial b_o} \quad \frac{\partial J_\mu}{\partial b_1} \quad \ldots \quad \frac{\partial J_\mu}{\partial b_{q-1}} \right|$$

is different from zero. From implicit function theorems they have solutions $b_f = B_f(b)$, f = 0, 1, ..., q-1, with continuous derivatives near b = 0 and with initial values $B_f(0) = 0$. The admissible one-parameter family of arcs is obtained from $y_i(x, b_o, b_1, \ldots, b_q)$ by replacing b_q by b and b_f by $B_f(b)$. This family contains E for b = 0, and when b is sufficiently small the arcs of this family belong to M. Replace the set of variations ξ_{cq}, $\eta_{iq}(x)$ by the set ξ_c, $\eta_i(x)$, then the variations along E of the one-parameter family are given by

$$\xi_{cf} B_f'(0) + \xi_c,\ \eta_{if}(x)\ B_f'(0) + \eta_i(x),$$

where the primes indicate differentiation with respect to b. If the n variations

$$\eta_{if}(x)\ B_f'(0) + \eta_i(x)$$

are not all identically zero, then the family will contain arcs not identical with E. Now when the functions η_{if} have been chosen to secure rank q for the matrix

$$\left\| \hat{J}_\mu \left(\mathcal{f}_{co}, \eta_{io} \right) \quad \cdots \quad \hat{J}_\mu \left(\mathcal{f}_{c\ q-1}, \eta_{i\ q-1} \right) \right\|$$

the variations η_i can always be selected linearly independent of them, thereby insuring that the variations

$$\eta_{if}(x)\ B_f'(0) + \eta_i(x)$$

are not identically zero. This selection can be made by determining the functions $\mathcal{f}_{if}(x)$ corresponding to the variations $\eta_{if}(x)$ by means of the equations

$$\overset{I}{\phi}{}^a_\beta = 0, \quad \overset{I}{\phi}{}^a_\gamma = \mathcal{f}_\gamma(x), \ \gamma = m_a + 1, \ \ldots, \ n,$$

and then selecting $\mathcal{f}_i(x)$ linearly independent of $\mathcal{f}_{if}$ $(f = 0,1,\ldots,$ q-1) and finally choosing for the variations $\eta_i(x)$ solutions of the equations

$$\overset{I}{\phi}{}^a_\beta (x, \eta, \eta') = 0, \qquad \overset{I}{\phi}{}^a_\gamma (x, \eta, \eta') = \mathcal{f}_\gamma(x),$$

with the functions $\mathcal{f}_i(x)$ substituted in these equations.

Corollary 4. If $\mathcal{f}_c, \eta_i(x)$ is a set of admissible variations satisfying the equations of variation $\overset{I}{\phi}{}^a_\beta = 0$, $\hat{J}_\mu = 0$ along a normal arc E ϵ M, then the one-parameter family of arcs in M imbedding E of theorem 4 can be so chosen that it has the set $\mathcal{f}_c, \eta_i(x)$ as its variations along E.

Proof. The one-parameter family constructed in Theorem 4 will suffice for this corollary provided $B_f'(0)$ appearing in the variations, all vanish. Consider the equations

$$J_\mu(B_f(b), b) = 0, \ \mu = 1, \ \ldots, \ q.$$

If we differentiate these equations, we have at b = 0 the equations

$$\hat{J}_\mu \left(\xi_{cf}, \eta_{if} \right) B_f'(0) + \hat{J}_\mu \left(\xi_c, \eta_i \right) = 0.$$

Since $\xi_c, \eta_i(x)$ satisfy the equations $\hat{J}_\mu = 0$, the above equation reduces to

$$\hat{J}_\mu \left(\xi_{cf}, \eta_{if} \right) B_f'(0) = 0.$$

Now the determinant

$$\left| \hat{J}_\mu \left(\xi_{cf}, \eta_{if} \right) \right|$$

has rank q and hence is different from zero. Hence the $B_f'(0)$ are zero.

We can now state another corollary to Theorem 4, and, because it is concerned with what happens on sub-arcs between corners of E, it is precisely the same result that Bliss obtained for the Bolza problem [1, 215]. We state it here without proof.

Corollary 5. Each of the sequence of elementary families which together form the one-parameter family of arcs in M described in Theorem 4 and Corollary 4, is defined by functions

$$y_i (x, b), \qquad x' \leq x \leq x'', \qquad |b| < \varepsilon,$$

for which the derivatives y_{ib}, y'_{ib} exist and are continuous in a neighborhood of values (x, b) satisfying the conditions $x' \leq x \leq x''$, $b = 0$. If the imbedded arc E is an extremal, so that the functions $y_i(x)$ defining it have continuous second derivatives, then the derivatives $y_{ibb}, y'_{ibb}, (y_{ibb})'$ also exist at the values (x, b) satisfying $x' \leq x \leq x''$, $b = 0$, and $y'_{ibb} = (y_{ibb})'$. On each elementary family the following differentials exist and satisfy the equations

$$dx_o = x_{ob}\, db, \ \ldots, \ dx_p = x_{pb}\, db, \ dy_i = y'_i\, db + \delta y_i,$$

$$d^2 y_i = y'_i\, d^2 x + y'_i\, dx^2 + 2\delta y'_i\, dx + \delta^2 y_i.$$

WEIERSTRASS CONDITION

We are now able to state and prove an analogue of the necessary condition of Weierstrass.

Theorem 5. An admissible arc E satisfying the equations $\varphi^a_\beta = 0$ and the multiplier rule, with multipliers $\lambda_0 = 1$, $\lambda_\beta(x)$, is said to satisfy this analogue of the necessary condition of Weierstrass with these multipliers if the condition

$$W^a(x, y, y', \lambda, Y') = F^a(x, y, Y', \lambda) - F^a(x, y, y', \lambda) - (Y'_i - y'_i) F^a_{y'_i}(x, y, y', \lambda) \geq 0$$

is valid at every element (x, y, y', λ) of E, except possibly at partition points of E, for all admissible sets $(x, y, Y') \neq (x, y, y')$ satisfying the equations $\varphi^a_\beta = 0$. Every normal minimizing arc E for this problem must satisfy this condition.

We need the following lemma in order to prove Theorem 5.

Lemma 2. Let E be a normal minimizing arc. Then there is a set of admissible variations $\xi_{of}, \xi_{1f}, \ldots, \xi_{pf}, \eta_{if}(x)$, $f = 0,1,\ldots, q-1$, satisfying the equations of variation $\bar{\phi}^a_\beta = 0$ along E such that

$$\left| \hat{J}_\mu (\xi_{cf}, \eta_{if}) \right| \neq 0.$$

Proof of Lemma 2. Suppose that for every set of q admissible variations satisfying $\bar{\phi}^a_\beta = 0$ the determinant $\left| \hat{J}_\mu(\xi, \eta) \right| = 0$. We consider the equation $\lambda_0 \hat{J} + e_\mu \hat{J}_\mu = 0$ which must be satisfied by every set of admissible variations. The condition that $\left| \hat{J}_\mu \right| = 0$ for a set of q admissible variations implies that $\lambda_0 = 0$, contradicting the normality of E.

Proof of Theorem 5. Let $\underline{a}$ be arbitrary and consider the stage associated with the interval $[x_{a-1}, x_a]$. As in the Bolza problem [1, 220], let t be an arbitrary point between corners of E^a. Let Y_i', $i = 1, 2, \ldots, n$, be a set of values such that the element $[x_t, y(x_t), Y']$ is admissible and satisfies the equations $\varphi_\beta^a = 0$. This system of differential equations can be enlarged as in the proof of Theorem 2 so that the continuity properties described there hold near the element $[x_t, y(x_t), Y']$ as well as near E^a with $\left|\varphi^a_{\alpha y_i'}\right| \neq 0$ at $[x_t, y(x_t), Y']$ as well as on E^a. The enlarged system defines a set of functions $z_\gamma(x)$ corresponding to the functions $y_i(x)$ defining E^a, and a set of constants Z_γ associated with the set $[x_t, y(x_t), Y']$. The equations of variation define functions $\mathcal{J}_{\gamma\sigma}(x)$, $\sigma = 1, \ldots, s$, corresponding to each of the sets of admissible variations $\xi_{c\sigma}$, $\eta_{i\sigma}(x)$ of Lemma 2. As in the Bolza problem, we can infer the existence of three families of admissible arcs

$$y_i(x, b), \qquad x_{a-1} - \delta < x \leq x_t, \qquad |b| < \varepsilon,$$

$$Y_i(x, b), \qquad x_t \leq x \leq x_t + e, \qquad |b| < \varepsilon, \qquad |e| < \varepsilon,$$

$$y_i(x, b, e) \qquad x_t + e \leq x < x_a + \delta, \qquad |b| < \varepsilon, \qquad |e| < \varepsilon,$$

satisfying differential equations

$$y_i' = M_i\,(x, y, z(x) + b_\sigma \mathcal{J}_\sigma), \qquad x_{a-1} - \delta \leq x \leq x_t,$$

$$y_i' = M_i\,(x, y, z), \qquad x_t \leq x \leq x_t + e,$$

$$y_i' = M_i\,(x, y, z(x) + b_\sigma \mathcal{J}_\sigma), \qquad x_t + e \leq x \leq x_a + \delta$$

and initial conditions

$$y_i\,(x_{a-1},\ b) = y_i\,(x_{a-1}) + b_\sigma\,\eta_{i\sigma}\,(x_{a-1}),$$

$$Y_i\,(x_t,\ b) = y_i\,(x_t,\ b),$$

$$y_i\,(x_t + e,\ b,\ e) = Y_i\,(x_t + e,\ b).$$

The system of differential equations is equivalent to

$$\varphi^a_\beta = 0, \quad \varphi^a_\gamma = Z_\gamma(x) + b_\sigma\,\eta_{i\sigma}\,(x), \qquad x_{a-1} - \delta < x \leq x_t,$$

$$\varphi^a_\beta = 0, \quad \varphi^a_\gamma = Z_\gamma, \qquad x_t \leq x \leq x_t + e,$$

$$\varphi^a_\beta = 0, \quad \varphi^a_\gamma = Z_\gamma(x) + b_\sigma\,\eta_{i\sigma}(x), \qquad x_t + e \leq x < x_a + \delta$$

with $Y_i' = M_i(x_t,\ y(x_t),\ Z)$. For values $e > 0$ the three families form a single admissible s-parameter family of arcs consisting of a finite sequence of adjacent elementary families. The functions defining these elementary families and their derivatives with respect to x have continuous partial derivatives with respect to the parameters b and e. Continuity with respect to b follows from the arguments of Theorem 2 and for e from well-known existence theorems in differential equations [1, 278].

If $b = e = 0$ then the first and third families reduce to the functions $y_i(x)$ defining the arc E^a. The variations along E^a with respect to b_σ of the first and third families satisfy the differential equations of variation (for the enlarged system) with the functions $\mathcal{f}_{i\sigma}(x)$ corresponding to the variations $\eta_{i\sigma}(x)$. If $\eta_i(x)$ denotes the variations along the arc E^a of the first and third families with respect to e, then they satisfy the equations

$$\overset{1}{\phi}{}^a_\beta = 0$$

and the relations

$$\eta_i(x) \equiv 0, \qquad x_{a-1} - \delta < x \leq x_t,$$

$$y_i'(x_t) + \eta_i(x_t) = Y_i'.$$

On each of the other stages one gets an s-parameter family $y(x, b)$ of comparison arcs and by matching up parameters these stage-wise comparison arcs piece together a family of comparison arcs for the problem under consideration. Furthermore, when

$$x_c(b) = x_c + b_\sigma \xi_{c\sigma}$$

are used to define staging points and substituted in the end and intermediate point constraints J_μ, then the functions J_μ become functions $J_\mu(b, e)$ of the parameters b, e. At the values $(b, e) = (0, 0)$ these functions have, as in the case of the Bolza problem, derivatives

$$\frac{\partial J_\mu}{\partial b_\sigma} = \hat{J}_\mu(\xi_\sigma, \eta_\sigma), \qquad \frac{\partial J_\mu}{\partial e} = \hat{J}_\mu(\xi, \eta).$$

The equations $J_\mu(b, e) = 0$ have initial solutions $(b, e) = (0, 0)$ at which functional determinant

$$\left| \frac{\partial J_\mu}{\partial b_\sigma} \right| = \left| \hat{J}_\mu(\xi_\sigma, \eta_\sigma) \right|$$

is different from zero. Only the second subscript on ξ_σ, η_σ is being used, actually these should read $\xi_{c\sigma}$, $\eta_{i\sigma}$. Now the equations $J_\mu(b, e) = 0$ have solutions $b_\sigma = B_\sigma(e)$ which vanish at $e = 0$ and have continuous derivatives near $e = 0$. These derivatives satisfy the equations

$$\hat{J}_\mu(\xi_{c\sigma}, \eta_{i\sigma}) B_\sigma'(0) + \hat{J}_\mu(\xi_c, \eta_i) = 0$$

at $e = 0$. By replacing the parameters b_σ by $b_\sigma = B_\sigma(e)$ in the comparison arcs and end and intermediate conditions, we get a one-parameter family

of comparison arcs which contain the minimizing arc E for e = 0 and which are admissible for sufficiently small positive values of e.

Now the function J can be written as a function of the parameters b, e as follows,

$$J(b,e) = g[x_o(b),\ldots,x_p(b),y(x_o(b),b),\ldots,y(x_a(b),b,e),\ldots,y(x_p(b),b)]$$

$$+\int_{x_o(b)}^{x_1(b)} f^1(x,y(x,b),y'(x,b))dx+\ldots+\int_{x_{a-2}(b)}^{x_{a-1}(b)} f^{a-1}(x,y(x,b),y'(x,b))dx$$

$$+\int_{x_{a-1}(b)}^{x_t} f^a(x,y(x,b),y'(x,b))dx+\int_{x_t}^{x_t+e} f^a(x,Y,Y')dx$$

$$+\int_{x_t+e}^{x_a(b)} f^a(x,y(x,b,e),y'(x,b,e))dx+\int_{x_a(b)}^{x_{a+1}(b)} f^{a+1}(x,y(x,b),y'(x,b))dx$$

$$+\ \ldots\ +\int_{x_{p-1}(b)}^{x_p(b)} f^p(x,y(x,b),y'(x,b))dx.$$

using precisely the same techniques as are used in the Bolza problem we find that the derivatives of J at (b, e) = (0, 0) are defined by

$$\frac{\partial J}{\partial b_\sigma} + e_\mu \hat{J}_\mu (\xi_{c\sigma}, \eta_{i\sigma}) = 0,$$

$$\frac{\partial J}{\partial e} + e_\mu \hat{J}_\mu (\xi_c, \eta_i) = W^a(x, y, y', \lambda , Y') \Big|_t,$$

where

$$W^a(x, y, y', \lambda , Y') = F^a(x, y, Y', \lambda) - F^a(x, y, y', \lambda)$$
$$- (Y_i' - y_i') \ F^a_{y_i'}(x, y, y', \lambda).$$

The arcs defined by functions y(x, b, e) for $x \in [x_{a-1}, x_a]$ are not admissible for small negative values of e but are admissible for

small positive values of e. Since J(E) is to be a minimum, as e increases from zero, the sum $J(B_\sigma(e), e)$ must be non-decreasing. Thus the derivative at e = 0 must be non-negative. The derivative of this sum at e = 0 is given by

$$J_{b_\sigma}(0, 0)\, B_\sigma'(0) + J_e(0, 0).$$

It follows easily then that $W^a(x, y, y', \lambda, Y') \geq 0$ between corners of E^a. One can also see from simple continuity arguments that

$$W^a(x, y, y', \lambda, Y') \geq 0$$

at corners of E^a.

CLEBSCH CONDITION

We follow the analogue of the Weierstrass condition by an analogue of the Clebsch condition.

Theorem 6. An admissible arc E satisfying the equations $\varphi^a_\beta = 0$ and the multiplier rule with multipliers $\lambda_0 = 1$, $\lambda_\beta(x)$ is said to satisfy this analogue of the necessary condition of Clebsch with these multipliers if the condition

$$F^a_{y_i' y_k'}(x, y, y', \lambda)\, \pi_i \pi_k \geq 0$$

holds at every element (x, y, y', λ) of E, except possibly at staging point, for all sets $(\pi_1, \pi_2, \ldots, \pi_n) \neq (0, 0, \ldots, 0)$ satisfying the equations $\varphi^a_{\beta y_i'}(x, y, y', \lambda)\, \pi_i = 0$. Every normal minimizing arc for this problem must satisfy this condition.

Proof. Let E be a normal minimizing arc for this problem. Let $\underline{a}$ be arbitrary and π_i, i = 1, 2, ..., n, be a set of values satisfying the equations

$$\varphi^a_{\beta y_i'}(x, y, y', \lambda)\,\pi_i = 0$$

where the element (x, y, y', λ) belongs to E. Now $n - m_a$ further quantities k_γ are defined by the equations

$$\varphi^a_{\gamma y_i'}(x, y, y', \lambda)\,\pi_i = k_\gamma.$$

The equations

$$\varphi^a_\beta(x, y, p) = 0, \qquad \varphi^a_\gamma(x, y, p) = z_\gamma + \varepsilon k_\gamma$$

have the initial solution (ε, p) and determine a set of solutions $p_i(\varepsilon)$ with initial values $p_i(0) = y_i'$. Now the above equations become

$$\varphi^a_\beta(x, y, p(\varepsilon)) = 0, \qquad \varphi^a_\gamma(x, y, p(\varepsilon)) = z_\gamma + \varepsilon k_\gamma$$

and differentiating with respect to ε and replacing $p_i(0)$ by y_i' in notation produces

$$\varphi^a_{\beta y_i'}\,p_i'(0) = 0, \qquad \varphi^a_{\gamma y_i'}\,p_i'(0) = k_\gamma,$$

so $p_i'(0) = \pi_i$. Now sets $(x, y, p(\varepsilon))$ are interior to R for sufficiently small ε, hence from Theorem 5, we have

$$W^a(x, y, y', \lambda, p(\varepsilon)) \geq 0.$$

Recall $W^a(x, y, y', \lambda, p(\varepsilon)) = F^a(x, y, p(\varepsilon), \lambda) - F^a(x, y, y', \lambda)$

$$- (p_i(\varepsilon) - y_i')\,F^a_{y_i'}(x, y, y', \lambda),$$

and note $W^a = 0$ at $\varepsilon = 0$ giving a minimum value to W^a. Differentiating with respect to ε and evaluating at $\varepsilon = 0$ produces

$$(E^a(0))' = \pi_i F^a_{y_i'} - \pi_i F^a_{y_i'} = 0$$

and $$(E^a(0))'' = F^a_{y_i' y_k'}\,\pi_i \pi_k.$$

Clearly, $F^a_{y_i' y_k'}\,\pi_i \pi_k$ must be non-negative and this completes the proof.

REFERENCES

1. Bliss, G. A., Lectures on the Calculus of Variations, The University of Chicago Press, Chicago, 1946.

2. Boyce, M. G. and Linnstaedter, J. L., "Necessary Conditions for a Multistage Bolza-Mayer Problem Involving Control Variables and Having Inequality and Finite Equation Constraints", Progress Report No. 7 on Studies in the Fields of Space Flight and Guidance Theory, TMX-53292, July 1965, NASA, pp. 8-31.

3. Denbow, C. H., "A Generalized Form of the Problem of Bolza", Contributions to the Calculus of Variations 1933-37, The University of Chicago Press, Chicago, 1937, pp. 449-484.

4. Hestenes, M. R., "A General Problem in the Calculus of Variations with Applications to Paths of Least Time", The Rand Corporation Research Memorandum RM-100, Santa Monica, California, March, 1950.

5. Miner, W. E. and Andrus, J. F., "Necessary Conditions for Optimal Lunar Trajectories with Discontinuous State Variables and Intermediate Point Constraints", AIAA Journal, Vol. 6, No. 11, Nov. 1968, pp. 2154-2159.

6. Valentine, F. A., "The Problem of Lagrange with Differential Equalities as Added Side Conditions", Contributions to the Calculus of Variations 1933-37, The University of Chicago Press, Chicago, 1937, pp. 403-447.

RELAXED CONTROLS FOR FUNCTIONAL EQUATIONS

By J. Warga
Professor, Northeastern University
Boston, Massachusetts

NASA Grant NGR 22-011-020
Supplement 1

RELAXED CONTROLS FOR FUNCTIONAL EQUATIONS*

by

J. Warga

1. Introduction. We wish to study a class of variational problems defined by functional equations and, in particular, by nonlinear integral equations. Special problems of this kind, involving one-dimensional "hereditary" and delay-differential equations were investigated, among others, by A. Friedman [1], M. N. Oğuztöreli [2], and A. Halanay [3] (see also [2] and [3] for other references to work on such one-dimensional problems). Control problems defined by multi-dimensional integral equations were discussed in a heuristic manner by A. G. Butkovskii [4]. The "usual" control problems, defined by ordinary differential equations, also represent a special case.

Among possible applications of our results, as specialized to integral equations, we may mention, in particular, nonlinear control problems defined by partial differential equations that are equivalent to Uryson integral equations. The methods that we employ are closely related to those previously developed in [5] and [6].

*This research was supported by N.A.S.A. Grant NGR 22-011-020, Supplement 1.

RELAXED CONTROLS FOR FUNCTIONAL EQUATIONS

As a convenient framework for our study we consider the following problem: let $\mathcal{Y}$ and Q be given spaces, $\mathcal{Y}$ Hausdorff, $\mathcal{U}$ a subset of Q, E_m the euclidean m-space, B_1 a closed subset of E_m, and $F: \mathcal{Y} \times Q \to \mathcal{Y}$ and $c = (c^1,..,c^m): \mathcal{Y} \times Q \to E_m$ given functions. The "original problem" consists in determining an "original minimizing point", that is, a point $(\bar{y},\bar{u}) \in \mathcal{Y} \times \mathcal{U}$ that minimizes $c^1(y,u)$ on the set $\{(y,u) \in \mathcal{Y} \times \mathcal{U} \mid y = F(y,u),\ c(y,u) \in B_1\}$; the "relaxed problem" consists in determining a "relaxed minimizing point" $(\bar{y},\bar{q})$ and an "approximate minimizing solution" $\{(y_i,u_i)\}_{i=1}^{\infty}$, that is, a point $(\bar{y},\bar{q}) \in \mathcal{Y} \times Q$ that minimizes $c^1(y,q)$ on the set $\{(y,q) \in \mathcal{Y} \times Q \mid y = F(y,q),\ c(y,q) \in B_1\}$, and a sequence $\{(y_i,u_i)\}_{i=1}^{\infty}$ in $\mathcal{Y} \times \mathcal{U}$ such that $y_i = F(y_i, u_i)$ and $\lim_{i\to\infty} c(y_i,u_i) = c(\bar{y},\bar{q})$.

This formulation is motivated by a typical model of a control problem: the parameter u describes the control functions and parameters (that can be chosen from some "admissible" set $\mathcal{U}$), the point y describes a motion of the system consistent with the chosen controls and subject to the "equation of motion"

$$(1.1) \quad y = F(y,u),$$

the relation

$$(1.2) \quad c(y,u) \in B_1$$

describes the restrictions imposed on the system, and c^1 is the cost functional.

RELAXED CONTROLS FOR FUNCTIONAL EQUATIONS

In general, as in the special case of variational problems defined by ordinary differential equations, the original problem , with controls in $\mathcal{U}$, does not admit a minimizing solution even if the functions F and c are "nice". We therefore embed $\mathcal{U}$ in a set Q of "relaxed controls" and define an appropriate topology on Q in which $\mathcal{U}$ is a dense subset of sequentially compact Q and F and c are continuous when restricted to the set $\{(y,q) \in \mathcal{Y} \times Q \mid y = F(y,q)\}$. This insures, subject to certain mild assumptions about F and c , the existence of a relaxed minimizing point $(\bar{y}, \bar{q})$ and of an approximate minimizing solution. The desired "relaxed" behavior of the system can be simulated by using an element of an approximate minimizing solution.

In studying necessary conditions for minimum we require somewhat different assumptions related to the nature of $\mathcal{Y}$ as a Banach space, the convexity of Q , the existence of (Frechet) derivatives F_y and c_y , and the invertibility of $I - F_y(y,q)$ at the relaxed minimizing point.

We observe, in §§3 and 4, that the usual optimal control problems defined by ordinary differential equations belong to the class of problems that we have described; so do the more general control problems defined by Uryson-type integral

equations that we discuss in some detail in §§3,4,6, and 7.

We discuss, in §2 , the following aspect of the general problem: (1) the existence of a relaxed minimizing point $(\bar{y},\bar{q}) \in \mathcal{Y} \times Q$; (2) the existence of an approximate minimizing solution; and (3) necessary conditions for a relaxed minimum. The corresponding proof is presented in §5 . We then apply these results in §§3 and 4 (with the proofs in §§6 and 7) to a control problem defined by a Uryson-type integral equation.

The general results for the Uryson-type relaxed control problem that we present in §§3 and 4 require rather complicated assumptions and setting that are introduced with the view toward generality and possible applications. As a consequence, the theorems are rather involved and the assumptions complicated. We therefore present, at first, less general results that have the advantage of greater simplicity.

1.3. The simplified Uryson-type control problem. Let T and R be compact subsets of some finite-dimensional euclidean spaces, $d\tau$ the Lebesgue measure on T, and $(t,\tau,v,r) \to g(t,\tau,v,r)$: $T \times T \times E_n \times R \to E_n$ continuous and such that g^i are independent of t for $i = 1,2,..,m \leq n$. We represent by S the class of regular Borel probability measures on R . The "original problem"

consists in determining functions $\bar{y}\colon\ T \to E_n$ and $\bar{\rho}\colon\ T \to R$ that yield the minimum of

$$c^1(y,\rho) = y^1 = \int_T g^1(\tau,y(\tau),\rho(\tau))d\tau$$

among all couples (y,ρ) for which y is continuous, ρ measurable,

$$(1.3.1) \qquad y(t) = \int_T g(t,\tau,y(\tau),\rho(\tau))d\tau \qquad (t \in T),$$

and

$$(1.3.2) \qquad y^i = 0 \qquad (i = 2,\ldots,m).$$

(Note that y^i are constant for $i \leq m$ since the corresponding g^i are independent of t). The "relaxed problem" consists in determining a relaxed minimizing solution $(\bar{y},\bar{\sigma})$, that is, functions $\bar{y}\colon\ T \to E_n$ and $\bar{\sigma}\colon\ T \to S$ that yield the minimum of

$$c^1(y,\sigma) = y^1 = \int_T d\tau \int_R g^1(\tau,y(\tau),r)\sigma(dr;\tau)$$

in the class $\mathcal{A}$ of all (y,σ) for which y is continuous, the function $\tau \to \int_R \phi(r)\sigma(dr;\tau)$ measurable for all continuous scalar ϕ,

$$(1.3.3) \qquad y(t) = \int_T d\tau \int_R g(t,\tau,y(\tau),r)\sigma(dr;\tau) \qquad (t \in T),$$

and

$$(1.3.4) \qquad y^i = 0 \qquad (i = 2,..,m).$$

We can state, as a consequence of the results presented in §§3 and 4, the following theorem:

Theorem 1.3.5. Assume that g and $g_v = (\partial g^i/\partial v^j)(i,j=1,..,n)$ exist and are uniformly continuous and bounded on $T \times T \times E_n \times R$, and that the class $\mathcal{A}$ is nonempty. Then there exists a relaxed minimizing solution $(\bar{y},\bar{\sigma})$.

If $\bar{y}$ is the unique continuous solution of the integral equation (1.3.3) for $\sigma = \bar{\sigma}$ then there exists a sequence $\{\rho_j\}_{j=1}^{\infty}$ of measurable functions and a sequence $\{y_j\}_{j=1}^{\infty}$ of continuous functions such that the (y_j,ρ_j) satisfy equation (1.3.1) for $j = 1,2,...$ and $\lim_{j\to\infty} y_j^i = \bar{y}^i$ for $i = 1,2,..,m$.

If the linear integral equation

$$w(t) = \int_T k(t,\tau)w(\tau)d\tau \qquad (t \in T)$$

has only the trivial solution $w(\cdot) = 0$ for

$k(t,\tau) = \int_R g_v(t,\tau,\bar{y}(\tau),r)\bar{\sigma}(dr;\tau)$ $(t,\tau \in T)$ then the relaxed minimizing solution $(\bar{y},\bar{\sigma})$ satisfies the following necessary condition for minimum:

there exist a nonvanishing $\hat{\lambda} = (\lambda^1,..,\lambda^m,0,..,0) \in E_n$ and a resolvent kernel $k^* = (k^{*i}_j)(i,j = 1,..,n)$ of k such that $(t,\tau) \to k^{*i}_j(t,\tau) = k^{*i}_j(\tau)$ are independent of t for $i = 1, .., m$ and

(Weierstrass E-condition or maximum principle)

$$\int_T d\tau \int_R \zeta(\tau) \cdot g(\tau,\vartheta,\bar{y}(\vartheta),r)\bar{\sigma}(dr;\vartheta)$$

$$= \operatorname*{Min}_{r' \in R} \int_T \zeta(\tau) \cdot g(\tau,\vartheta,\bar{y}(\vartheta),r')d\tau \quad \text{for almost all } \vartheta \in T,$$

where

$$\zeta(\tau) = (\zeta^1(\tau),\ldots,\zeta^n(\tau))$$

and

$$\zeta^j(\tau) = \sum_{i=1}^{m} \hat{\lambda}^i k^{*i}{}_j(\tau) + \hat{\lambda}^j/|T| \qquad (j = 1,..,n;\ \tau \in T).$$

(We say, in the present context, that k^* is a resolvent kernel of k if the equations

$$w(t) = \int_T k(t,\tau)w(\tau)d\tau + h(t) \quad \text{and} \quad w(t) = \int_T k^*(t,\tau)h(\tau)d\tau + h(t) \quad (t \in T)$$

are equivalent for continuous w and h).

The above theorem, which we prove in §8, is much too weak for our purposes: it does not even apply to control problems defined by ordinary differential equations. We consider, therefore, in §§3 and 4, a Uryson-type control problem in a more general setting: the sets T and R are assumed metric and compact, the "original controls" ρ are restricted by the condition

$$\rho(t) \in R^{\#}(t) \quad (t \in T) \quad (\text{where } t \to R^{\#}(t) \subset R \text{ is given})$$

and the "relaxed controls" σ satisfy analogous restrictions, the

function g has, as an additional argument, a "control parameter" b in a metric and compact space B, the condition $(y^2,..,y^n) = 0$ is replaced by $(y^1,..,y^n) \in B_1$ for a given B_1, the uniform continuity and boundedness of g and g_y are replaced by weaker assumptions, and the class $\mathcal{Y}$ of solutions y of the integral equation is extended beyond the class of continuous functions. We then study the existence aspects of the control problem for integral equations assuming $\mathcal{Y}$ to be $L^1(T,E_n)$; and we examine necessary conditions for a relaxed minimum assuming that $\mathcal{Y}$ is either $L^p(T,E_n)$ for $1 < p < \infty$ or $C(T,E_n)$. Necessary conditions for an original minimum will be discussed separately along the lines of [6]. We might mention, finally, that certain more general unilateral and minimax control problems that have been investigated for ordinary differential equations [7], [8], [9] extend quite naturally to integral equations; but we have only partial results so far.

I wish to acknowledge with thanks several stimulating conversations with J. Frampton.

2. The General Control Problem. Lemmas 2.1 and 2.2 below are obvious and are stated only to motivate the corresponding statements concerning the Uryson-type control problems and their proofs (Theorems 3.2 and 3.3). Theorem 2.3, on necessary conditions for minimum, is patterned after [6, Theorem 2.2, p. 644] and relies ultimately on a construction of McShane [10, pp. 17-18].

RELAXED CONTROLS FOR FUNCTIONAL EQUATIONS

We use the term "derivative" to mean "Frechet derivative" and the notation $h_x(x_1,y_1)$, $h_y(x_1,y_1)$ to represent partial derivatives. If h is defined for x in a subset Γ of a Banach space $\mathcal{X}$ and y in a Banach space $\mathcal{Y}$, with values in a Banach space $\mathcal{Z}$, we say that h has a derivative $h_{(x,y)}(x_1,y_1)$ at (x_1,y_1) relative to $\Gamma \times \mathcal{Y}$ if $h_{(x,y)}(x_1,y_1)$ is a linear operator from $\mathcal{X} \times \mathcal{Y}$ to $\mathcal{Z}$ such that

$$|h(x_2,y_2) - h(x_1,y_1) - h_{(x,y)}(x_1,y_1)((x_2,y_2) - (x_1,y_1))| =$$

$$o(|x_2 - x_1|_{\mathcal{X}} + |y_2 - y_1|_{\mathcal{Y}}) \quad \text{for all } x_2 \in \Gamma \text{ and}$$

for all $y_2 \in \mathcal{Y}$. The symbol I represents the identity operator on $\mathcal{Y}$. If Q is a convex subset of a linear space, $\mathcal{X}$ is any set and h is a function from $\mathcal{X} \times Q$ to some Banach space, we write $Dh(x,\bar{q};q - \bar{q})$ for $\lim_{\alpha \to +0} \frac{1}{\alpha}(h(x,\bar{q} + \alpha(q - \bar{q})) - h(x,\bar{q}))$.

We denote by $\bar{A}$ the closure of A.

Lemma 2.1 Let $\mathcal{Y}$ and Q be Hausdorff spaces, Q and $Y_1 = \{y \in \mathcal{Y} \mid y = F(y,q),\ c(y,q) \in B_1,\ q \in Q\}$ sequentially compact and F and c continuous when restricted to $\bar{Y}_1 \times Q$. Then either there exists a point $(\bar{y},\bar{q})$ that minimizes $c^1(y,q)$ on $\{(y,q) \in \mathcal{Y} \times Q \mid y = F(y,q),\ c(y,q) \in B_1\}$, or that set is empty.

Lemma 2.2 Let $\mathcal{Y}$ and Q be Hausdorff spaces, $\mathcal{U}$ a

dense subset of Q, and $(\bar{y}, \bar{q})$ a relaxed minimizing solution. Assume, furthermore, that Q satisfies the first axiom of countability and that

(2.2.1) $\bar{y}$ is the unique solution of the equation $y = F(y, \bar{q})$;

(2.2.2) there exists a neighborhood $\tilde{Q}$ of $\bar{q}$ such that the equation $y = F(y,q)$ has at least one solution y for each $q \in \mathcal{U} \cap \tilde{Q}$; and

(2.2.3) the set $Y_2 = \{y \in \mathcal{Y} \mid y = F(y,q), q \in \tilde{Q}\}$ is sequentially compact and F and c are continuous when restriced to $\bar{Y}_2 \times \tilde{Q}$.

Then there exists an approximate minimizing solution.

Theorem 2.3. Let $\mathcal{Y}$ be a Banach space, Q a convex subset of a linear space, $\omega^{\square}$ an array with real elements ω^{ij} $(i,j = 1,\ldots,m)$ considered as an element of E_{m^2} with origin $0^{\square}$, $\Omega = \{\omega^{\square} \mid \omega^{ij} \geq 0, \sum_{i,j=1}^{m} \omega^{ij} \leq 1\}$, and $(\bar{y},\bar{q})$ a relaxed minimizing point. Assume, furthermore, that for each fixed

subset $\{q_{ij} \mid i, j = 1,\ldots,m\}$ of Q there exists a neighborhood $\tilde{\mathcal{Y}} \times \tilde{\Omega}$ of $(\bar{y}, 0^{\Pi})$ in $\mathcal{Y} \times \Omega$ such that the functions $(y, \omega^{\Pi}) \to F(y, \bar{q} + \sum_{i,j=1}^{m} \omega^{ij}(q_{ij} - \bar{q})) : \tilde{\mathcal{Y}} \times \tilde{\Omega} \to \mathcal{Y}$ and

$(y, \omega^{\Pi}) \to c(y, \bar{q} + \sum_{i,j=1}^{m} \omega^{ij}(q_{ij} - \bar{q})) : \tilde{\mathcal{Y}} \times \tilde{\Omega} \to E_m$ are continuous, have derivatives at $(\bar{y}, 0^{\Pi})$ (relative to $\mathcal{Y} \times \tilde{\Omega}$) and continuous partial derivatives with respect to y on $\tilde{\mathcal{Y}} \times \tilde{\Omega}$, and that the operator $I - F_y(\bar{y}, \bar{q})$ is a linear homeomorphism of $\mathcal{Y}$ onto $\mathcal{Y}$. Let K_1 be a convex set in some E_ℓ, $\bar{\xi} \in K_1$, and $\phi = (\phi^1,\ldots,\phi^m) : K_1 \to B_1$ a continuous mapping with a derivative at $\bar{\xi}$ and such that $\phi(\bar{\xi}) = c(\bar{y}, \bar{q})$. Then either

$$(2.3.1)\quad \phi^1_\xi(\bar{\xi})\bar{\xi} = \operatorname*{Min}_{\xi \in K_1} \phi^1_\xi(\bar{\xi})\xi,$$

or there exist $\gamma \geq 0$ and $\lambda \in E_m$ such that $|\lambda| \neq 0$,

$$(2.3.2)\quad \lambda \cdot \{c_y(\bar{y},\bar{q})(I - F_y(\bar{y},\bar{q}))^{-1} DF(\bar{y},\bar{q};\, q - \bar{q}) + Dc(\bar{y},\bar{q};\, q - \bar{q})\} \geq 0 \quad \text{for all } q \in Q,$$

and

$$(2.3.3)\quad (\gamma\, \delta_1 - \lambda)\, \phi_\xi(\bar{\xi})\bar{\xi} = \operatorname*{Min}_{\xi \in K_1} (\gamma \delta_1 - \lambda)\, \phi_\xi(\bar{\xi})\xi,$$

where $\delta_1 = (1,0,\ldots,0) \in E_m$.

3. Control Problems Defined by Uryson-type Integral Equations. Existence of Relaxed and Approximate Minimizing Solutions. Let T, R and B be compact metric spaces. We assume that a nonnegative, finite, regular, complete, and nonatomic measure dt is defined on the Lebesgue extension of the Borel field of sets in T and we consider the corresponding

product measure $dtd\tau$ on $T \times T$. The symbol $|M|$ represents the measure of $M \subset T$, $\int h(t)dt$ the integral over T, $|a,b|$ the distance in a metric space, and $|a|$ (or $|a|_E$) the norm in a normed linear space E. We represent by $L^p(T,\mathcal{X})$ $(1 \leq p < \infty)$ the Banach space of measurable functions h from T to a Banach space $\mathcal{X}$ with the norm $|h(\cdot)|_p = \{\int |h(t)|^p_{\mathcal{X}}\, dt\}^{1/p}$ and by $C(T,\mathcal{X})$ the Banach space of continuous h from T to $\mathcal{X}$ with the norm $|h(\cdot)|_\infty = \sup_{t\in T} |h(t)|_{\mathcal{X}}$. We also set $L^p(T) = L^p(T, E_1)$ and $C(T) = C(T,E_1)$.

Original and relaxed controls.

Let $\mathcal{K}$ be the class of measurable mappings from T to R. As in [5], we refer to functions from T to R as "original controls". Let S be the class of regular Borel probability measures on R, and $\mathcal{J}$ the class of "measurable relaxed controls", that is, mappings σ from T to S that are measurable in the sense that $t \to \int_R \phi(r)\, \sigma(dr; t)$ is measurable on T for every continuous $\phi : R \to E_1$. We define $\mathcal{K}$ as a subset of $\mathcal{J}$ by identifying the function $t \to \rho(t)$ with the function $t \to \sigma_\rho(t)$, where $\sigma_\rho(t)$ is a measure concentrated at $\rho(t)$ with mass 1. We also identify all controls, original or relaxed, that differ only on sets of measure 0.

Topology in the space of measurable relaxed controls.

We define a topology in $\mathcal{J}$ as in [5, p.631]; we represent by $\mathcal{B}$ the Banach space (which is actually the space $L^1(T,C(R))$) of real-valued functions ϕ on $T \times R$, continuous on R for every t, measurable on T for every r, with $t \to \sup_{r\in R} |\phi(t,r)|$ integrable, and with $|\phi|_{\mathcal{B}} = \int \sup_{r\in R} |\phi(t,r)|\, dt$. We then define every $\sigma \in \mathcal{J}$ as an element of $\mathcal{B}^*$ (the topological dual of $\mathcal{B}$) by setting $\langle \sigma, \phi \rangle = \int dt \int_R \phi(t,r)\, \sigma(dr;t)$

for all $\phi \in \mathcal{B}$. The topology we choose for $\mathcal{B}^*$, and its subsets $\mathcal{S}$ and $\mathcal{R}$, is the weak star topology in $\mathcal{B}^*$ (the $\mathcal{B}$ topology of $\mathcal{B}^*$). It follows that $\lim_{i\to\infty} \sigma_i = \sigma$ implies $\lim_{i\to\infty} \int dt \int_R \phi(t,r)\sigma_i(dr;t)$ $\int dt \int_R \phi(t,r)\sigma(dr;t)$ for every $\phi \in \mathcal{B}$.

Sets $\mathcal{R}^\#$ and $\mathcal{S}^\#$ of restricted controls.

For a given mapping $R^\#$ from T to the class of nonempty subsets of R, we set $\mathcal{R}^\# = \{\rho \in \mathcal{R} \mid \rho(t) \in R^\#(t) \text{ on } T\}$ and $\mathcal{S}^\# = \{\sigma \in \mathcal{S} \mid \sigma(\bar{R}^\#(t),t) = 1 \text{ on } T\}$, where $\bar{R}^\#(t)$ is the closure of $R^\#(t)$.

We shall consider mappings $R^\#$ satisfying either or both of the following assumptions ([5, Assumption 2.3, p. 631]):

(3.1.1) For every $\epsilon > 0$ there exists a closed subset T_ϵ of T, of measure at least $|T| - \epsilon$, such that for every $\bar{t} \in T_\epsilon$ and every $r \in \bar{R}^\#(\bar{t})$ there exists an original control $\rho \in \mathcal{R}^\#$, continuous at $\bar{t}$ when restricted to T_ϵ, and such that $|\rho(\bar{t}), r| < \epsilon$.

(3.1.2) For every $\epsilon > 0$ there exists a closed subset T_ϵ of T, of measure at least $|T| - \epsilon$, such that the mapping $R^\#$, when restricted to T_ϵ, is continuous with respect to the Hausdorff distance of sets $|R^\#(t_1), R^\#(t_2)|$ (where, for $A, B \subset R$, $|A,B| = \inf \{h \mid A \subset U(B,h), B \subset U(A,h)\}$ and $U(A,h)$ is the h-neighborhood of A in R). Here we identify all subsets of R whose mutual Hausdorff distance is 0.

Formulation of the Uryson-type control problem.

Now let $g = (g^1,\ldots,g^n)$, and let $(t,\tau,v,r,b) \to g(t,\tau,v,r,b)$ $T \times T \times E_n \times R \times B \to E_n$ be measurable in (t,τ) for every fixed

(v,r,b) and continuous in (v,r,b) for every fixed (t,τ). We also assume that $g^i(t,\tau,v,r,b) = g^i(\tau,v,r,b)$ $(i = 1,\ldots,m \leq n)$ is independent of t for all (τ,v,r,b). Let

$$f(t,\tau,v,s,b) = \int_R g(t,\tau,v,r,b)\, s(dr)$$

for all (t, τ, v, b) and all $s \in S$. We consider solutions (y, σ, b) of the integral equation

$$y(t) = \int f(t, \tau, y(\tau), \sigma(\tau), b)d\tau \qquad (t \in T)$$

in $\mathcal{Y} \times \mathcal{A}^{\#} \times B$, where $\mathcal{Y}$ is some Banach space of measurable functions from T to E_n.

A solution (y, σ, b) is "a relaxed admissible solution" if $(y^1, y^2,\ldots,y^m) \in B_1$ (observe that $y^i (i = 1,\ldots,m)$ are constant on T since $g^i(t, \tau, v, r, b)$ $(i = 1,\ldots,m)$ are independent of t). A relaxed admissible solution $(\bar{y}, \bar{\sigma}, \bar{b})$ is "a relaxed minimizing solution" if $\bar{y}^1 \leq y^1$ for all relaxed admissible solutions (y, σ, b).

We relate the control problem just described to the general problem discussed in §2 in the following manner: let $\mathcal{U} = \mathcal{K}^{\#} \times B$ and $Q = \mathcal{A}^{\#} \times B$. We let the mappings $(y, q) \to F(y, q) = F(y, \sigma, b)$ and $(y, q) \to c(y, q) = c(y, \sigma, b)$ be defined, for $q = (\sigma, b) \in \mathcal{A}^{\#} \times B$ and $y \in \mathcal{Y}$, by

$$F(y, \sigma, b)(t) = \int f(t, \tau, y(\tau), \sigma(\tau), b)d\tau \qquad (t \in T),$$

$$c^i(y, \sigma, b) = \int f^i(\tau, y(\tau), \sigma(\tau), b)d\tau \qquad (i = 1,\ldots,m),$$

if this defines $F(y,\sigma,b)$ as an element of $\mathcal{Y}$. Otherwise we set, for some $a \in \mathcal{Y}$, $a \neq 0$,

$$F(y, \sigma, b) = y + a,$$

$$c(y, \sigma, b) = 0.$$

We can easily verify that, in the case where T is the interval $[t_0, t_1]$ of the real axis and $g(t, \tau, v, r, b)$ has a special form, the Uryson integral equation becomes an ordinary differential equation, our control problem the "standard" control problem, and the results that follow a slight generalization of previous results [5, Theorem 3.1, p. 633], [6, Theorem 3.4, p. 648]. We further discuss this problem in § 4.

We can now state existence and approximation theorems that we prove in §5. In both of these theorems we set $\mathcal{Y} = L^1(T, E_n)$.

Theorem 3.2 There exists a relaxed minimizing solution if the following conditions are satisfied:

(3.2.1) the class of relaxed admissible solutions is nonempty for $\mathcal{Y} = L^1(T, E_n)$;

(3.2.2) $R^{\#}$ satisfies Assumption (3.1.2);

and either

(3.2.3) there exists a positive function ψ, integrable on $T \times T$ and such that, for every solution (y, σ, b) of the equation $y = F(y,\sigma,b)$, we have

$$|g(t, \tau, y(\tau), r, b)| \leq \psi(t, \tau) \quad \text{on} \quad T \times T \times R \times B,$$

or

(3.2.4) there exist real numbers c_1, p, and β and a measurable ψ_0 on $T \times T$ such that $0 \leq \beta < \infty$, $1 \leq p \leq \infty$, $p \geq \beta$,

$$|g(t, \tau, v, r, b)| \leq (1 + |v|^{\beta}) \psi_0(t,\tau) \quad \text{on} \quad T \times T \times E_n \times R \times$$

$$\int |\psi_0(t, \cdot)|_{p/(p-\beta)}^{p} \, dt < \infty,$$

and every solution (y, σ, b) of the equation $y = F(y, \sigma, b)$ is such that $|y(\cdot)|_p \leq c_1$.

Theorem 3.3 Let $R^{\#}$ satisfy Assumptions (3.1.1) and (3.1.2), and let $\bar{y}$ b the unique solution of the equation $y = F(y,\sigma,b)$ for $\sigma = \bar{\sigma} \in \mathcal{S}^{\#}$ and $b = \bar{b} \in B$. Assume, furthermore, that

(3.3.1) the equation $y = F(y, \rho, \bar{b})$ admits at least one solution y in $L^1(T, E_n)$ for each $\rho \in \mathcal{R}^{\#}$ in some neighborhood of $\bar{\sigma}$, and either condition (3.2.3) or condition (3.2.4) is satisfied. Then there exists a sequence $\{\rho_i\}_{i=1}^{\infty}$ in $\mathcal{R}^{\#}$ and a sequence $\{y_i\}_{i=1}^{\infty}$ in $L^1(T, E_n)$ such that $y_i = F(y_i, \rho_i, \bar{b})$ and $\lim_{i\to\infty} c(y_i, \rho_i, \bar{b}) = c(\bar{y}, \bar{\sigma}, \bar{b})$.

4. Necessary Conditions for a Relaxed Minimum of a Uryson-type Control Problem. We shall investigate necessary conditions for a point $(\bar{y}, \bar{\sigma}, \bar{b})$ to be a relaxed minimizing solution in a somewhat different framework than was required in §3.

Assumption 4.1

(4.1.1) $\mathcal{Y}$ is either $L^p(T, E_n)$ for $1 < p < \infty$ or $C(T, E_n)$, and T and R have the properties described in §3;

(4.1.2) B is a convex subset of a linear space;

(4.1.3) for every fixed choice of $b_{ij} \in B$ $(i, j = 1,\ldots,m)$ there exists $\omega_{max} \in (0, 1/m^2]$ such that, for $\Omega = \{\omega^{\square} = (\omega^{ij})(i, j = 1,\ldots,m) \mid 0 \leq \omega^{ij} \leq \omega_{max}\} \subset E_{m^2}$, the function $(t,\tau,v,r,\omega^{\square}) \to g^{\#}(t,\tau,v,r,\omega^{\square}) = g(t,\tau,v,r,\bar{b} + \sum_{i,j=1}^{m} \omega^{ij}(b_{ij}-\bar{b}))$: $T \times T \times E_n \times R \times \Omega \to E_n$ has a derivative with respect to $(v,\omega^{\square})$, and $g^{\#}$, $g_v^{\#}$, and $g_{\omega^{\square}}^{\#}$ are measurable in (t,τ)

for every $(v, r, \omega^{\square})$ and continuous in $(v, r, \omega^{\square})$ for every (t,τ);

(4.1.4) if $\mathcal{Y} = L^p(T, E_n)$ then there exist measurable positive functions ψ_0 and ψ_1 on $T \times T$ and numbers α and β such that $0 \le \alpha \le p-1$, $0 \le \beta \le p$, $\int |\psi_0(\cdot,\tau)|_p^{p/(p-\beta)}\, d\tau < \infty$, $\int |\psi_1(t,\cdot)|_{p/(p-1-\alpha)}^{p}\, dt < \infty$, $\int |\psi_1(\cdot,\tau)|_p^{p/(p-1-\alpha)}\, d\tau < \infty$, and, for all $(t, \tau, v, r) \in T \times T \times E_n \times R$, $b_{ij} \in B$, and $\omega^{\square} \in \Omega$,

$$|g^{\#}(t, \tau, v, r, \omega^{\square})| \le (1 + |v|^{\beta})\ \psi_0(t,\tau),$$

$$|g^{\#}_{v}(t, \tau, v, r, \omega^{\square})| \le (1 + |v|^{\alpha})\ \psi_1(t, \tau),$$

and

$$|g^{\#}_{\omega^{\square}}(t, \tau, v, r, \omega^{\square})| \le (1 + |v|^{\beta})\ \psi_0(t, \tau);$$

if $\mathcal{Y} = C(T, E_n)$ then there exists a compact set D in E_n containing $\{\bar{y}(t) \mid t \in T\}$ in its interior and integrable ψ_0 and ψ_1 on T such that

$$|g^{\#}(t, \tau, v, r, \omega^{\square})| \le \psi_0(\tau),$$

$$|g^{\#}_{v}(t, \tau, v, r, \omega^{\square})| \le \psi_1(\tau),$$

and

$$|g^{\#}_{\omega^{\square}}(t, \tau, v, r, \omega^{\square}) \le \psi_0(\tau)$$

for all $(t, \tau, v, r) \in T \times T \times D \times R$, $b_{ij} \in B$, and $\omega^{\square} \in \Omega$. Furthermore, there exists a positive function $h \to \Phi(h)$ such that, for $t_1, t_2 \in T$ and $\hat{g} = g^{\#}, g^{\#}_{v}$, and $g^{\#}_{\omega^{\square}}$,

$$\int \sup_{D\times R\times\Omega} | \hat{g}(t_1, \tau, v, r, \omega^0) - \hat{g}(t_2, \tau, v, r, \omega^0)| d\tau \leq \Phi(|t_1, t_2|)$$

and $\lim_{h \to +0} \Phi(h) = 0$;

(4.1.5) for $k(t,\tau) = f_v(t, \tau, \bar{y}(\tau), \bar{\sigma}(\tau), \bar{b})$ on $T \times T$, the integral equation

$$w(t) = \int k(t, \tau)\, w(\tau) d\tau \qquad (t \in T)$$

has only the solution $w(\cdot) = 0$ in $\mathcal{Y}$.

Resolvent kernel. If there exists a measurable real matrix-valued function $k^* = (k^{*i}_j)(i, j = 1,\ldots,n)$ on $T \times T$ such that, for every $x \in \mathcal{Y}$, the relations

$$w(t) = \int k(t, \tau) w(\tau) d\tau + x(t) \qquad \text{a.e. in } T$$

and

$$w(t) = \int k^*(t, \tau) x(\tau) d\tau + x(t) \qquad \text{a.e. in } T$$

are equivalent in $\mathcal{Y}$, we refer to k^* as a resolvent kernel of k.

We can now state necessary conditions for a relaxed minimum in Theorem 4.2 below. Conditions (1), (2), and (3) of (4.2.2) are generalizations of respectively the Weierstrass E-condition, transversality with respect to parameters and initial conditions, and transversality with respect to the end conditions of the calculus of variations.

Theorem 4.2 Let $(\bar{y}, \bar{\sigma}, \bar{b})$ be a relaxed minimizing solution, and let Assumption 4.1 be satisfied. Let $\mathcal{R}^{\#}_{\infty}$ be a

denumerable subset of $\mathcal{L}^{\#}$, $R^*(t) = \{\rho(t) \mid \rho \in \mathcal{L}^{\#}_{\infty}\}$ $(t \in T)$, $\overline{R}^*(t)$ the closure of $R^*(t)$, $S^*(t) = \{s \in S \mid s(\overline{R}^*(t)) = 1\}$, K_1 a convex set in some E_ℓ, $\overline{\xi} \in K_1$, and $\phi : K_1 \to B_1$ a continuous mapping with a derivative at $\overline{\xi}$ and such that $\phi^i(\overline{\xi}) = c^i(\overline{y}, \overline{\sigma}, \overline{b}) = \overline{y}^{\,i}$ $(i = 1,\ldots,m)$. Then

(4.2.1) there exists a resolvent kernel k^* of k such that k^{*i}_j is independent of t for $i = 1,2,\ldots,m$; and $\int |k^*(\cdot,\tau)|^{p/(p-1)}_p d\tau < \infty$ if $\mathcal{Y} = L^p(T, E_n)$ and $\int \sup_{t \in T} |k^*(t,\tau)| d\tau < \infty$ if $\mathcal{Y} = C(T, E_n)$;

(4.2.2) either $\phi^1_\xi(\overline{\xi})\overline{\xi} = \operatorname{Min}_{\xi \in K_1} \phi^1_\xi(\overline{\xi})\xi$, or there exist a nonvanishing $\lambda = (\lambda^1,\ldots,\lambda^m) \in E_m$ and $\gamma \geq 0$ such that, setting

$$\hat{\lambda} = (\lambda^1,\ldots,\lambda^m,0,\ldots 0) = (\lambda,0,\ldots,0) \in E_n,$$

$$\zeta^j(\tau) = \sum_{i=1}^{m} \lambda^i k^{*i}_j(\tau) + \frac{\hat{\lambda}^j}{|T|} \quad (\tau \in T,\ j=1,\ldots,n),$$

$$\zeta(\tau) = (\zeta^1(\tau),\ldots,\zeta^n(\tau)) \quad (\tau \in T),$$

$$H_1(s,\theta) = \int \zeta(\tau) \cdot f(\tau, \theta, \overline{y}(\theta), s, \overline{b}) d\tau \qquad ((s,\theta) \in S \times T),$$

and

$$H_2(b) = \int_{T \times T} \zeta(\tau) \cdot Df(\tau,\theta, \overline{y}(\theta), \overline{\sigma}(\theta), \overline{b}; b - \overline{b}) d\tau d\theta \quad (b \in B),$$

the following conditions are satisfied:

(The Weierstrass E-condition)

(1) $H_1(\overline{\sigma}(\theta),\theta) = \operatorname{Min}_{s \in S^*(\theta)} H_1(s,\theta) =$

$$\operatorname{Min}_{r \in \overline{R}^*(\theta)} \int \zeta(\tau) \cdot g(\tau, \theta, \overline{y}(\theta), r, \overline{b}) d\tau \quad \text{for almost all } \theta \in T,$$

(Transversality Conditions)

(2) $\operatorname{Min}_{b \in B} H_2(b) = 0,$

and

$$(3)\quad (\gamma\delta_1 - \lambda)\phi_\xi(\bar{\xi})\bar{\xi} = \operatorname*{Min}_{\xi\in K_1} (\gamma\delta_1 - \lambda)\phi_\xi(\bar{\xi})\xi, \quad \text{where}$$

$$\delta_1 = (1, 0,\ldots,0) \in E_m.$$

In particular, if $R^{\#}(t) = R$ for all $t \in T$, $\bar{R}^*(\theta)$ and $S^*(\theta)$ can be replaced by R and S, respectively, in relation (1).

An illustration. As an illustration, we shall apply Theorem 4.2 to the following "standard" relaxed control problem: let T be the closed interval $[t_0,t_1]$ of the real axis, $d\tau$ the Lebesgue measure on T, $B_1 \subset E_m$, B a convex subset of some E_ℓ, ϕ_0 a continuously differentiable mapping from B into E_m with the image B_0, and h: $T \times E_m \times R \to E_m$. We wish to determine functions $\bar{x}$: $T \to E_m$ and $\bar{\sigma}$: $T \to S$ that yield the minimum of $x^1(t_1)$ among all absolutely continuous x and all measurable (in the previously defined sense) σ that satisfy the relations:

$$\frac{dx(\tau)}{d\tau} = \int_R h(\tau, x(\tau), r)\sigma(dr;\tau) \quad \text{a.e. in } T,$$

$$x(t_0) \in B_0,\ x(t_1) \in B_1.$$

We set $n = 2m$, $g = (g^1,..,g^n)$, $y = (y^1,..,y^n)$, $v = (w_1,w_2)$ with $w_1,w_2 \in E_m$, and, for all $(t,\tau,v,r,b) \in T \times T \times E_n \times R \times B$ and $i = 1,2,..,m$,

$$y^{i+m}(t) = x^i(t),$$

$$y^i(t) = x^i(t_1),$$

$$g^{i+m}(t,\tau,v,r,b) = \begin{cases} h(\tau,w_2,r) + \phi_0(b)/(t_1-t_0) & \text{for } \tau \le t, \\ \phi_0(b)/(t_1-t_0) & \text{for } \tau > t, \end{cases}$$

$$g^i(t,\tau,v,r,b) = g^{i+m}(t_1,\tau,v,r,b).$$

We then observe that our new problem is formally equivalent to the Uryson-type control problem considered in §3 and in the present section. We can easily verify that Theorem 4.2 is applicable if we set $\mathcal{Y} = C(T,E_n)$ and assume that

$$(\tau,w,r) \to h(\tau,w,r) \quad \text{and} \quad h_w(\tau,w,r)$$

exist on $T \times E_m \times R$, are continuous in (w,r) and measurable in τ, and that $|h(\tau,w,r)|$ and $|h_w(\tau,w,r)|$ are bounded by an integrable function of τ for all $(w,r) \in D \times R$, where D is some compact set in E_m containing the trajectory $\{\bar{x}(t) \mid t \in T\}$ in its interior.

We can evaluate the resolvent kernel k^* by a straightforward (if somewhat tedious) computation and determine that

$$\zeta^j(\tau) = \frac{1}{t_1-t_0}\lambda^j = \frac{1}{t_1-t_0} z^j(t_1) \quad (\tau \in T,\ j = 1,..,m),$$

$$\zeta^{j+m}(\tau) = -dz^j(\tau)/d\tau \quad (\tau \in T,\ j = 1,..,m),$$

where the absolutely continuous function $\tau \to z(\tau)\colon\ T \to E_m$ is the solution of the system

$$\frac{dz(\tau)}{d\tau} = \dot{z}(\tau) = - A^T(\tau)z(\tau) \qquad \text{a.e. in } T,$$

$$z(t_1) = \lambda,$$

A^T is the transpose of A, and $A(\tau) = \int_R h_w(\tau,\bar{x}(\tau),r)\bar{\sigma}(dr;\tau)$ $(\tau \varepsilon T)$. It follows then that

$$H_1(\vartheta,r) = z(\vartheta)\cdot \int_R h(\vartheta,\bar{x}(\vartheta),r)\bar{\sigma}(dr;\vartheta) + \frac{1}{|T|} z(t_0)\cdot\bar{x}(t_0) \quad (\vartheta \varepsilon T, r \varepsilon R)$$

Thus relation (1) of Theorem 4.2 yields the familiar Weierstrass E-condition (maximum principle) for the relaxed control problem defined by ordinary differential equations. In a similar manner, relations (2) and (3) yield the support (transversality) conditions at the initial and terminal points t_0 and t_1, respectively.

5. Proofs of the Statements in §2. The proofs of Lemmas 2.1 and 2.2 are trivial and will be omitted.

Proof of Theorem 2.3 We first consider the special case where $B_1 = \{(b^1, \ldots, b^m) \mid b^i = 0 \ (i = 2, \ldots, m)\}$.

Consider the equation

$$(5.1) \quad y = F(y, \bar{q} + \sum_{i,j=1}^{m} \omega^{ij}(q_{ij} - \bar{q}))$$

for an arbitrary choice of $q^{\square} = (q_{ij})$ with $q_{ij} \in Q$. We can apply, with minor changes, the proof of the implicit function theorem [II, p. 265] to show that there exists a neighborhood $\tilde{Y} \times \tilde{\Omega}$ of $(\bar{y}, 0^{\square})$ relative to $\mathcal{Y} \times \Omega$ such that equation (5.1) has a unique solution $y = \eta(\omega^{\square}, q^{\square}) \in \tilde{Y}$ for every $\omega^{\square} \in \tilde{\Omega}$ and the function $\omega^{\square} \to \eta(\omega^{\square}, q^{\square}) : \tilde{\Omega} \to \tilde{Y}$ is continuous and has a derivative at $0^{\square}$. It follows that

$$\omega^{\square} \to \tilde{c}(\omega^{\square}, q^{\square}) = c(\eta(\omega^{\square}, q^{\square}), \bar{q} + \sum_{i,j=1}^{m} \omega^{ij}(q_{ij} - \bar{q})) : \tilde{\Omega} \to E_m$$

is also continuous and has a derivative at $0^{\square}$.

Now let $\theta^{11} = \theta$, $\theta^{ij} = 0$ $((i,j) \neq (1,1))$, $\tilde{q}_{ij} = q(i,j = 1,..,m)$, $h(\theta; q) = \tilde{c}(\theta^{\square}, \tilde{q}^{\square})$, $V = \{dh(0; q)/d\theta \mid q \in Q\}$, and let W be the convex cone in E_m generated by V; that is, $W = \{\sum_{i=1}^{m} a^i v_i \mid v_i \in V, a^i \geq 0\}$. We shall show in the sequel that there exists $\lambda \in E_m$ such that $|\lambda| \neq 0$, $\lambda^1 \geq 0$, and $\lambda \cdot w \geq 0$ for all $w \in W$.

If this last statement is true, then $\lambda \cdot dh(0;q)/d\theta \geq 0$ for all $q \in Q$. We have $dh(0;q)/d\theta =$

$c_y(\bar{y},\bar{q})\eta_{\omega 11}(0^{\square}, \tilde{q}^{\square}) + Dc(\bar{y}, \bar{q}; q - \bar{q})$. Also, the differentiation of both sides of the equation $\eta(\theta^{\square}; \tilde{q}^{\square}) = F(\eta(\theta^{\square}; \tilde{q}^{\square}), \bar{q} + \theta^{11}(q - \bar{q}))$ with respect to θ^{11} at $0^{\square}$ yields

$$\eta_{\omega 11}(0^{\square}, \tilde{q}^{\square}) = F_y(\bar{y}, \bar{q})\eta_{\omega 11}(0^{\square}, \tilde{q}^{\square}) + DF(\bar{y}, \bar{q}; q - \bar{q}).$$

We then conclude that

$$(5.2)\quad \lambda \cdot dh(0;q)/d\theta = \lambda \cdot \{c_y(\bar{y},\bar{q})(I - F_y(\bar{y},\bar{q}))^{-1}DF(\bar{y},\bar{q};q - \bar{q}) + Dc(\bar{y},\bar{q}; q - \bar{q})\} \geq 0 \qquad \text{for all } q \in Q.$$

We now proceed to prove that there exists a point λ as just described. Indeed, assume the contrary. Then it follows from elementary properties of convex sets that there exists a point $w = (w^1, 0,....,0)$ in the interior of W, linearly independent $w_i \in W$ and positive $\alpha^i (i = 1,...,m)$ such that

$$w^1 < 0 \quad \text{and} \quad w = \sum_{i=1}^{m} \alpha^i w_i.$$

By the definition of W, there exist points q_{ij} and numbers

a^{ij} $(i, j = 1,\ldots,m)$ such that $a^{ij} \geq 0$ and

$$w_i = \sum_{j=1}^{m} a^{ij}\, dh(0; q_{ij})/d\theta \qquad (i = 1,\ldots,m).$$

Since the w_i are independent, the matrix $(w_i^j)(i, j = 1,\ldots,m)$ is nonsingular.

Now let $\tilde{\Gamma} = \{\gamma \in E_m \mid 0 \leq \gamma^i \leq \gamma_{max} (i = 1,\ldots,m)\}$, where γ_{max} is positive and sufficiently small so that $\omega^{\square}(\gamma) = (\omega^{ij}(\gamma)) = (\gamma^i a^{ij}) \in \tilde{\Omega}$, and consider the function $\gamma \to k(\gamma) = \tilde{c}(\omega^{\square}(\gamma); q^{\square}) : \tilde{\Gamma} \to E_m$. This function is continuous and has a derivative $k_\gamma(0) = (\partial k(0)/\partial\gamma^1,\ldots,\partial k(0)/\partial\gamma^m)$ at 0 relative to $\tilde{\Gamma}$ (where $\partial k(0)/\partial\gamma^i$ are right-hand derivatives). Furthermore, $\partial k(0)/\partial\gamma^{\ell} = \sum_{i,j=1}^{m} dh(0; q_{ij})/d\theta \cdot \partial\omega^{ij}(0)/\partial\gamma^{\ell} = \sum_{j=1}^{m} dh(0; q_{\ell j})/d\theta \cdot a^{\ell j}$

$= w_\ell$ $(\ell = 1,\ldots,m)$; hence the derivative $k_\gamma(0) = (w_i^j)$ has an inverse and $k_\gamma(0)\alpha = w = (w^1, 0,\ldots,0)$ for $\alpha = (\alpha^1,\ldots,\alpha^m)$. It follows (as in [**6**, p. 650]) that there exists a solution $\epsilon \to \gamma(\epsilon)$ of the equation

$$k(\gamma(\epsilon)) - k(\gamma(0)) = \epsilon w$$

for all sufficiently small positive ϵ, and $\gamma^i(\epsilon) \underset{\epsilon \to +0}{\to} +0$ $(i = 1,..,m)$. There exist, therefore, $q = \bar{q} + \sum_{i,j=1}^{m} \omega^{ij}(\gamma(\epsilon))(q_{ij} - \bar{q}) \in Q$ and $y = \eta(\omega^{\square}(\gamma(\epsilon)); q^{\square}) \in \}$ such that $y = F(y,q)$, $c^i(y,q) = 0$ $(i = 2,\ldots,m)$, and $c^1(y,q) < c^1(\bar{y}, \bar{q})$, contradicting the assumption that $(\bar{y},\bar{q})$ is a relaxed minimizing point.

We conclude that when $B_1 = \{(b^1,\ldots,b^m) \mid b^i = 0 (i = 2,\ldots,m)\}$ there exists a point $\lambda \in E_m$ such that $|\lambda| \neq 0$, $\lambda^1 \geq 0$, and relation (5.2) is satisfied.

We now consider the general case and define the sets $Q^{\#}$ and $B_1^{\#}$ and the functions $F^{\#} : \mathcal{Y} \times Q^{\#} \to \mathcal{Y}$ and $c^{\#}: \mathcal{Y} \times Q^{\#} \to E_{m+1}$ by

$$Q^{\#} = Q \times K_1, \quad B_1^{\#} = \{(v^0, v^1, \ldots, v^m) \mid v^i = 0 \ (i = 1, \ldots, m)\}$$

$$F^{\#}(y,q,\xi) = F(y,q),$$

$$c^{\#0}(y,q,\xi) = c^1(y,q),$$

$$c^{\#\ell}(y,q,\xi) = c^{\ell}(y,q) - \phi^{\ell}(\xi) \quad (\ell = 1, \ldots, m).$$

Clearly, the point $(\bar{y}, \bar{q}, \bar{\xi})$ is a relaxed minimizing point for the problem defined by $\mathcal{Y}$, $Q^{\#}$, $B_1^{\#}$, $F^{\#}$ and $c^{\#}$, which is of the form just investigated. The conclusions of the theorem follow from relation (5.2) applied to the transformed problem; the details of the argument are as in [6, Proof of Theorem 2.2, p. 650]. QED

6. Existence of relaxed and approximate minimizing solutions. for Uryson-type problems. Proofs.

Lemma 6.1 Let conditions (3.1.2) and (3.2.3) be satisfied, and let

$Y_2 = \{y \in \mathcal{Y} \mid y = F(y, \sigma, b), \ \sigma \in \mathcal{J}^{\#}, \ b \in B\}$. Then every sequence $\{y_i\}_{i=1}^{\infty}$ in Y_2 has a subsequence converging to some $\tilde{y} \in \mathcal{Y} = L^1(T, E_n)$.

Proof: Let $y_i = F(y_i, \sigma_i, b_i)$ $(i = 1, 2, \ldots,)$, $y_i \in \mathcal{Y}$, $\sigma_i \in \mathcal{J}^{\#}$, and $b_i \in B$. We must show that there exist a sequence J of natural numbers and a point $\tilde{y}$ in $\mathcal{Y}$ such that $\lim_{i \in J} y_i = \tilde{y}$ in $\mathcal{Y}$.

Let, for $v \in E_n$, $\chi(v) = 1$ if $|v| \leq 1$ and

$$\chi(v) = \frac{1}{|v|^2} \qquad \text{if} \quad |v| > 1,$$

$$\tilde{g}(t, \tau, v, r, b) = \chi(|g(t, \tau, v, r, b)|/\psi(t, \tau))g(t, \tau, v, r, b),$$

and

$$\tilde{f}(t, \tau, v, s, b) = \int_R \tilde{g}(t, \tau, v, r, b)s(dr),$$

for all t, τ, v, r, b and all $s \in S$. Then, by (3,2,3), $\tilde{g}(t, \tau, y(\tau), r, b) = g(t, \tau, y(\tau), r, b)$ on $T \times T \times R \times B$ for every $y \in Y_2$; hence every solution (y, σ, b) of the equation $y = F(y, \sigma, b)$ also satisfies the equation

$$(6.1.1) \quad y(t) = \int \tilde{f}(t, \tau, y(\tau), \sigma(\tau), b)d\tau \qquad (t \in T).$$

Furthermore,

$$|\tilde{g}(t, \tau, v, r, b)| \leq \psi(t, \tau) \quad \text{on} \quad T \times T \times E_n \times R \times B$$

and $\tilde{g}$ is continuous in (v, r, b) and measurable in (t, τ).

Now let $\tilde{\psi}(t) = \int\psi(t, \tau)d\tau$ on T, $S_N = \{v \in E_n | \ |v| \leq N\}$, $P_N = \{t \in T \mid \tilde{\psi}(t) \leq N\}$, and $\epsilon > 0$. Then there exists $N = N(\epsilon)$ such that, for $P = P_{N(\epsilon)}$,

$$(6.1.2) \quad \int dt\int_{T-P}\psi(t, \tau)d\tau \leq \frac{1}{8}\epsilon.$$

Since $\tilde{g}$ is measurable in (t, τ), continuous in (v, r, b) onthe compact set $S_N \times R \times B$, and $|\tilde{g}(t, \tau, v, r, b)| \leq \psi(t, \tau)$, the restriction of $\tilde{g}^i$ to $T \times P \times S_N \times R \times B$ is, for each $i = 1,\ldots,n$, an element of $L^1(T \times P, C(S_N \times R \times B))$; there exist, therefore, an integer $k = k(\epsilon)$ and functions $\alpha_j \in L^1(T \times P, E_n)$ and $\beta_j \in C(S_N \times R \times B)$ such that

$$(6.1.3)\quad \int_T\int_P \max_{S_N\times R\times B} \left| \tilde{g}(t,\tau,v,r,b) - \sum_{j=1}^{k} \beta_j(v,r,b)\,\alpha_j(t,\tau) \right| dtd\tau \le \tfrac{1}{4}\,\epsilon.$$

Furthermore, each $\alpha_j \in L^1(T \times P, E_n)$ can be approximated by a finite sum $\sum_{\ell} b_{\ell}(\tau)\, a_{\ell}(t)$, where b_{ℓ} are measurable characteristic functions on P and $a_{\ell} \in L^1(T, E_n)$. We conclude, therefore, that relation (6.1.3) can be rewritten, by appropriately changing the definitions of k and β_j, as

$$(6.1.4)\quad \int_T\int_P \max_{S_N\times R\times B} \left|\tilde{g}(t,\tau,v,r,b) - \sum_{j=1}^{k} \beta_j(v,r,b)b_j(\tau)a_j(t) \right| dtd\tau \le \tfrac{1}{4}\,\epsilon,$$

and we may also assume that $|\beta_j(v,r,b)| \le 1$ on $S_N\times R\times B$.

Now let

$$\gamma_{ji} = \gamma_{ji}(\epsilon) = \int_P b_j(\tau)d\tau \int_R \beta_j(y_i(\tau),r,b_i)\sigma_i(dr;\tau).$$

We observe that

$$|y_i(t)| \le \int |\tilde{f}(t,\tau,y_i(\tau),\sigma_i(\tau),b_i)|\, d\tau \le \int \psi(t,\tau)d\tau \le N = N(\epsilon)$$

for $t \in P = P(\epsilon)$ and all $i = 1, 2, 3....$ Therefore, for all integers p and q, for all $t \in T$, and for $k = k(\epsilon)$,

$$|y_p(t) - y_q(t)| \le 2\int_{T-P} \psi(t,\tau)d\tau + \Big|\int_P d\tau\Big\{\int_R \tilde{g}(t,\tau,y_p(\tau),r,b_p)\sigma_p(dr;\tau$$
$$- \int_R \tilde{g}(t,\tau,y_q(\tau),r,b_q)\sigma_q(dr$$
$$\le 2\int_{T-P} \psi(t,\tau)d\tau + \Big|\sum_{j=1}^{k} (\gamma_{jp} - \gamma_{jq})a_j(t)\Big|$$
$$+ \sum_{i=p,i=q} \int_P \max_{r\in R} \Big|\tilde{g}(t,\tau,y_i(\tau),r,b_i) - \sum_{j=1}^{k} \beta_j(y_i(\tau),r,b_i)b_j(\tau)a_j(\tau)\Big|d\tau;$$

hence, in view of relations (6.1.2) and (6.1.4),

$$(6.1.5)\quad \int_T |y_p(t) - y_q(t)|dt \le \tfrac{1}{4}\,\epsilon + \sum_{j=1}^{k} |\gamma_{jp} - \gamma_{jq}| \int_T |a_j(t)|dt + \tfrac{1}{2}\,\epsilon.$$

Given any infinite subsequence $\tilde{J}$ of $\{1, 2, \ldots\}$, we can determine a subsequence $J' = J'(\tilde{J}, \epsilon)$ such that the sequences $\{\gamma_{ji}(\epsilon)\}_{i \in J'}$ have a limit, for each $j = 1, 2, \ldots k(\epsilon)$, since $|\gamma_{ji}| \leq |P| \leq |T|$ for all i and j. Now let $J_0 = \{1, 2, \ldots\}$, $J_{\ell+1} = J'(J_\ell, 2^{-\ell})$ $(\ell = 0, 1, 2, \ldots)$, and let $\bar{J}$ be the diagonal subsequence of $J_0, J_1, \ldots$. Then $\{\gamma_{ji}(\epsilon)\}_{i \in \bar{J}}$ converges for each $\epsilon > 0$ and $j = 1, 2, \ldots, k(\epsilon)$, and there exists an integer $i_0 = i_0(\epsilon)$ such that, in view of (6.1.5),

$$\int_T |y_p(t) - y_q(t)|\, dt \leq \epsilon$$

provided $p \geq q \geq i_0(\epsilon)$ and $p, q \in \bar{J}$.

We conclude that $\{y_i(\cdot)\}_{i \in \bar{J}}$ is a Cauchy sequence in $L^1(T, E_n)$ and converges, therefore, to some $\tilde{y}$ in $L^1(T, E_n)$. QED

Lemma 6.2 Condition (3.2.4) implies condition (3.2.3).

Proof. Let (y, σ, b) satisfy the equation $y = F(y, \sigma, b)$, and assume that condition (3.2.4) is satisfied. Then, for $t \in T$,

$$|y(t)| \leq \left|\int_T d\tau \int_R \tilde{g}(t, \tau, y(\tau), r, b)\sigma(dr; \tau)\right| \leq \int (1 + |y(\tau)|^\beta)\psi_0(t, \tau)\, d\tau$$

and, by Hölder's inequality,

$$|y(t)| \leq \int \psi_0(t, \tau)\, d\tau + |y(\cdot)|_p^\beta \, |\psi_0(t, \cdot)|_{p/(p-\beta)} \leq |\psi_0(t, \cdot)|_{p/(p-\beta)} |T|^{\cdot}$$

$$|\psi_0(t, \cdot)|_{p/(p-\beta)}\, c_1^\beta = \gamma |\psi_0(t, \cdot)|_{p/(p-\beta)}, \quad \text{where } \gamma = c_1^\beta + |T|^{\beta/p}.$$

It follows that, for all $(t, \tau, r, b) \in T \times T \times R \times B$,

$$|g(t, \tau, y(\tau), r, b)| \leq (1 + \gamma^\beta |\psi_0(\tau, \cdot)|_{p/(p-\beta)}^\beta)\psi_0(t, \tau)$$

Now let the expression on the right be denoted by $\psi(t,\tau)$, let $c_2 = \{\int |\psi_0(\tau,\cdot)|^p_{p/(p-\beta)}\, d\tau\}^{\beta/p}$ and let $\psi'(t,\tau) = \psi_0(t,\tau)|\psi_0(\tau,\cdot)|^{\beta}_{p/(p-\beta)}$. Then, by Hölder's inequality,

$$\int \psi'(t,\tau)d\tau \le |\psi_0(t,\cdot)|_{p/(p-\beta)} \{\int |\psi_0(t,\cdot)|^p_{p/(p-\beta)} dt\}^{\beta/p}$$

and

$$\int\int \psi'(t,\tau)dtd\tau \le |T|^{(p-1)/p}\, c_2^{1+\beta}.$$

Also $\int_{T\times T}\psi_0(t,\tau)dtd\tau < \infty$. It follows that $(t,\tau) \to \psi(t,\tau)$ is integrable on $T \times T$. QED

6.3 Proof of Theorem 3.2. Because of Lemma 6.2 it suffices to assume that conditions (3.2.1), (3.2.2) and (3.2.3) are satisfied. Now let $\{(y_i, \sigma_i, b_i)\}_{i=1}^{\infty}$ be a sequence in $\mathcal{Y} \times \mathcal{S}^{\#} \times B$ and $y_i = F(y_i, \sigma_i, b_i)$. By (3.2.2) and [6, Theorem 2.5, p. 632] the set $\mathcal{S}^{\#}$ is sequentially compact, and by Lemma 6.1 there exists a sequence J and a $\tilde{y} \in \mathcal{Y}$ such that $\lim_{i\in J} y_i = \tilde{y}$. We may choose J so that $\lim_{i\in J} \sigma_i = \tilde{\sigma}$ and $\lim_{i\in J} b_i = \tilde{b}$ for some $\tilde{\sigma} \in \mathcal{S}^{\#}$ and $\tilde{b} \in B$, and $\lim_{i\in J} y_i = \tilde{y}$ a.e. in T, say for $t \in T'$.

For each fixed t and τ in T', $g(t, \tau, \cdot, \cdot, \cdot)$ is continuous, hence uniformly continuous, on the compact set $D_\tau \times R \times B$, where D_τ is a compact subset of E_n containing $\tilde{y}(\tau)$ in its interior. It follows that $\lim_{i\in J} g(t,\tau,y_i(\tau),\cdot,b_i) = g(t,\tau,y(\tau),\cdot,\tilde{b})$ uniformly on R and

$$\lim_{i\in J} \int_R (g(t,\tau,y_i(\tau),r,b_i) - g(t,\tau,\tilde{y}(\tau),r,\tilde{b}))\sigma_i(dr;\tau) =$$

$\lim_{i\in J} \alpha_i(t,\tau) = 0$ for all $t,\tau \in T'$. Furthermore,

$|\alpha_i(t,\tau)| \leq 2\psi(t,\tau)$ on $T' \times T'$ and $\psi(t,\cdot)$ is integrable; therefore, for each $t \in T'$,

$$\tilde{y}(t) = \lim_{i \in J} y_i(t) =$$

$$\lim_{i \in J} \int_T d\tau \int_R (g(t,\tau,\tilde{y}_i(\tau), r, b_i) - g(t,\tau,\tilde{y}(\tau),r,\tilde{b}))\sigma_i(dr;\tau)$$

$$+ \lim_{i \in J} \int_T d\tau \int_R g(t,\tau,\tilde{y}(\tau),r,\tilde{b})\sigma_i(dr;\tau)$$

$$= \lim_{i \in J} \int_T d\tau \int_R g(t,\tau,\tilde{y}(\tau),r,\tilde{b})\sigma_i(dr;\tau)$$

$$= \int_T d\tau \int_R g(t,\tau,\tilde{y}(\tau),r,\tilde{b})\tilde{\sigma}(dr;\tau) = \int_T f(t,\tau,\tilde{y}(\tau),\tilde{\sigma}(t),\tilde{b})d\tau ,$$

since the function $(\tau,r) \to g(t,\tau,\tilde{y}(\tau),r,\tilde{b}) \in \mathcal{B}$ (as defined in §3).

Thus $\tilde{y}(t) = F(\tilde{y},\tilde{\sigma},\tilde{b})(t)$ for $t \in T'$. By redefining $\tilde{y}$, if necessary, on $T - T'$, we can assert that $\tilde{y} = F(\tilde{y},\tilde{\sigma},\tilde{b})$ and thus the set of solutions of $y = F(y,\sigma,b)$ is nonempty and sequentially compact in $\mathcal{Y} \times \mathcal{L}^{\#} \times B$. Since $y^j = c^j(y,\sigma,b)$ $(j = 1,\ldots,m)$ for every solution (y,σ,b), the $y^j (j = 1,\ldots,m)$ are constant, and B_1 is closed, it follows that there exists a minimizing relaxed solution. QED

6.4 Proof of Theorem 3.3. By Assumptions (3.1.1) and (3.1.2) and by [5, Theorem 2.4, p. 631], the set $\mathcal{K}^{\#}$ is a dense subset of $\mathcal{L}^{\#}$. There exists, therefore, a sequence $\{\rho_i\}_{i=1}^{\infty}$ in $\mathcal{K}^{\#}$ converging to $\bar{\sigma}$. By (3.3.1), there exists an integer i_0 and a sequence $\{y_i\}_{i=i_0}^{\infty}$ in $\mathcal{Y}$ such that $y_i = F(y_i, \rho_i, \bar{b})$. It follows then, as in the proof of Theorem 3.2, that there exist $\tilde{y} \in \mathcal{Y}$ and a sequence J such that $\lim_{i \in J} y_i = \tilde{y}$, $\lim_{i \in J} \rho_i = \bar{\sigma}$, and $\tilde{y} = F(\tilde{y},\bar{\sigma},\bar{b})$. By the uniqueness assumption, $\tilde{y} = \bar{y}$.

Thus $\lim_{i\in J} c^j(y_i, \rho_i, \bar{b}) = \lim y_i^j = \bar{y}^j$ $(j = 1,\ldots,m)$ since the y_i^j $(j = 1,\ldots,m)$ are constant. QED

7. Proof of Theorem 4.2. We shall use the notation and the assumptions of §4. We also set, for a fixed choice of b_{ij} $(i,j = 1,\ldots,m)$ in B and $\sigma_{ij}(i,j = 1,\ldots,m)$ in $\mathcal{S}^{\#}$, and for all $\omega^{\square} \in \Omega$, $v \in E_n$, $y \in \mathcal{Y}$, and $t \in T$,

$$\tilde{b}(\omega^{\square}) = \bar{b} + \sum_{i,j=1}^{m} \omega^{ij}(b_{ij} - \bar{b}),$$

$$\tilde{\sigma}(\omega^{\square}) = \bar{\sigma} + \sum_{i,j=1}^{m} \omega^{ij}(\sigma_{ij} - \bar{\sigma}),$$

$$\tilde{f}(t,\tau,v,\omega^{\square}) = f(t,\tau,v,\sigma(t;\omega^{\square}),\tilde{b}(\omega^{\square})),$$

$$\tilde{F}(y,\omega^{\square})(t) = \int \tilde{f}(t,\tau,y(\tau),\omega^{\square})d\tau .$$

Lemma 7.1 Let b_{ij} and $\sigma_{ij}(i,j = 1,\ldots,m)$ be fixed. Then $(y,\omega^{\square}) \to \tilde{F}(y,\omega^{\square})$: $\mathcal{Y} \times \Omega \to \mathcal{Y}$ is continuous in some neighborhood Γ of $(\bar{y},0^{\square})$ and has a derivative at $(\bar{y}, 0^{\square})$, the partial derivative $\tilde{F}_y$ exists and is continuous on Γ, and the following relations hold:

$$(\tilde{F}_y(y,\omega^{\square})\Delta y)(t) = \int_T \tilde{f}_v(t,\tau,y(\tau),\omega^{\square})\Delta y(\tau)d\tau \quad (t\in T,\ y\in\mathcal{Y},\Delta y\in\mathcal{Y},\ \omega^{\square}\in\Omega),$$

$$\tilde{F}_{\omega^{\square}}(\bar{y},0^{\square})(t) = \int_T \tilde{f}_{\omega^0}(t,\tau,y(\tau),0^{\square})d\tau \quad (t \in T,\ y \in \mathcal{Y}),$$

$$(\tilde{F}_y(\bar{y},0^{\square})\Delta y)(t) = \int_T k(t,\tau)\Delta y(\tau)d\tau \quad (t \in T,\ \Delta y \in \mathcal{Y}),$$

and

$$\tilde{F}_{\omega_{11}}(\bar{y},0^{\square})(t) = \int_T f(t,\tau,\bar{y}(\tau),\sigma_{11}(\tau) - \bar{\sigma}(\tau),\bar{b})d\tau +$$

$$\int_T Df(t,\tau,\bar{y}(\tau),\ \bar{\sigma}(\tau),\ \bar{b};\ b_{11} - \bar{b})d\tau \quad (t \in T).$$

Proof: We first consider the case $\mathcal{Y} = L^p(T, E_n)$ for $1 < p < \infty$. Let $y \in \mathcal{Y}$ and $\omega^{\square} \in \Omega$ be fixed. We observe that the function $(t, \tau) \to \tilde{f}(t,\tau,y(\tau),\omega^{\square})$ is measurable on $T \times T$ and $\int |\tilde{f}(t,\tau,y(\tau),\omega^{\square})|^p dt \leq$ $\int |\psi_0(t,\tau)|^p \ (1 + |y(\tau)|^{\beta})^p dt \leq |\psi_0(\cdot,\tau)|_p^p \ (1 + |y(\tau)|^{\beta})^p < \infty$ for almost all $\tau \in T$. Thus the function $t \to \tilde{f}(t,\tau,y(\tau),\omega^{\square})$ belongs to $L^p(T,E_n)$ for almost all $\tau \in T$ and [9, Lemma 16, p. 196] $\tau \to \tilde{f}(\cdot,\tau,y(\tau),\omega^{\square})$ is a measurable function from T to $L^p(T, E_n)$. Furthermore, $\tau \to 1 + |y(\tau)|^{\beta} \in L^{p/\beta}(T)$ and $\tau \to |\psi_0(\cdot,\tau)|_p \in L^{p/(p-\beta)}(T)$; hence, by Hölder's inequality,

$$\int |\tilde{f}(\cdot,\tau,y(\tau),\omega^{\square})|_p \, d\tau \leq \int |\psi_0(\cdot,\tau)|_p \ (1 + |y(\tau)|^{\beta}) d\tau < \infty.$$

Thus $\tau \to \tilde{f}(\cdot,\tau,y(\tau),\omega^{\square})$ is an integrable function from T to $L^p(T, E_n)$ for all $y \in L^p(T, E_n)$ and $\omega^{\square} \in \Omega$, and $\tilde{F}$ exists on $L^p(T, E_n) \times \Omega$.

Now consider the continuity of $\tilde{F}$ and the existence and continuity of $\tilde{F}_y$. We have, for fixed $y \in \mathcal{Y}$ and $\omega^{\square} \in \Omega$, and for all $\Delta y \in \mathcal{Y}$,

$$(7.1.1) \quad \tilde{F}(y + \Delta y,\omega^{\square})(t) - \tilde{F}(y,\omega^{\square})(t) =$$

$$\int \{\tilde{f}(t,\tau,y(\tau) + \Delta y(\tau), \omega^{\square}) - \tilde{f}(t,\tau,y(\tau),\omega^{\square})\} d\tau =$$

$$\int \tilde{f}_v(t,\tau,y(\tau) + \theta(t,\tau)\Delta y(\tau),\omega^{\square})\Delta y(\tau) d\tau \quad \text{a.e. in } T,$$

where $0 \leq \theta(t,\tau) \leq 1$, and we may assume (using essentially the argument in [13, Lemma 18.1, p. 177]) that $(t,\tau) \to \theta(t,\tau)$ is measurable. Furthermore,

$|\tilde{f}_v(t,\tau,y(\tau) + \theta(t,\tau)\Delta y(\tau),\omega^{\square})| \leq$

$(1 + (|y(\tau)| + |\Delta y(\tau)|)^{\alpha})\psi_1(t,\tau)$;

hence

$\int|\tilde{f}_v(t,\tau,y(\tau) + \theta(t,\tau)\Delta y(\tau),\omega^{\square})|^p dt \leq$

$(1 + (|y(\tau)| + |\Delta y(\tau)|)^{\alpha})^p|\psi_1(\cdot,\tau)|_p^p$.

It follows that $t \to \tilde{f}_v(t,\tau,y(\tau) + \theta(t,\tau)\Delta y(\tau),\omega^{\square})$ belongs to $L^p(T,E_{n^2})$ for almost all $\tau \in T$ and

$|\tilde{F}(y + \Delta y,\omega^{\square}) - \tilde{F}(y,\omega^{\square})|_p \leq$

$\int(1 + (|y(\tau)| + |\Delta y(\tau)|)^{\alpha})|\psi_1(\cdot,\tau)|_p\ |\Delta y(\tau)|\ d\tau$.

We can easily verify that, for a fixed y in $\mathcal{Y}$, the coefficient of $|\Delta y(\tau)|$ in the integrand on the right has an $L^{p/(p-1)}$ norm bounded by some constant c_1 for all Δy in the unit ball of $L^p(T,E_n)$.

We conclude that

$$|\tilde{F}(y + \Delta y,\omega^{\square}) - \tilde{F}(y,\omega^{\square})|_p \leq c_1|\Delta y|_p$$

for all $\Delta y \in L^p(T, E_n)$ and $\omega^{\square} \in \Omega$. Thus $y \to \tilde{F}(y,\omega^{\square})$ is continuous at every y, uniformly in $\omega^{\square} \in \Omega$.

Our previous argument shows that the function $\Delta y \to \mathcal{D}_1(\Delta y;\ y) = \int\tilde{f}_v(\cdot,\tau,y(\tau),\omega^{\square})\Delta y(\tau)d\tau$ is a bounded linear operator on $L^p(T,\ E_n)$ for every $(y,\omega^{\square})$. Relation (7.1.1) now yields

(7.1.2) $|\tilde{F}(y + \Delta y, \omega^{\square}) - \tilde{F}(y, \omega^{\square}) - \mathcal{O}_1(\Delta y; y)\rangle(t)| \leq$

$$\leq \int \sup_{R\times\Omega} |g_v^{\#}(t,\tau,y(\tau) + \theta(t,\tau)\Delta y(\tau), x, \omega^{\beta}) - g_v^{\#}(t,\tau,y(\tau), x, \omega^{\beta})| \; |\Delta y(\tau)| \, d\tau.$$

As Δy converges to 0 in $L^p(T, E_n)$, hence also in measure, the coefficient of $|\Delta y(\tau)|$ in the integrand on the right converges to 0 in measure, as a function of τ, for almost all $t \in T$. This coefficient is also bounded by $\alpha(t,\tau) = \psi_1(t,\tau)(2 + |y(\tau)|^{\alpha} + (|y(\tau)| + |\Delta y(\tau)|)^{\alpha})$, and we verify that $t \to |\alpha(t,\cdot)|_{p/(p-1)}$ belongs to $L^p(T)$. It follows, applying Hölder's inequality to the right side of (7.1.2) and then taking the L^p-norm with respect to t, that

$$\lim_{|\Delta y|_p \to 0} |\tilde{F}(y + \Delta y, \omega^{\square}) - \tilde{F}(y, \omega^{\square}) - \mathcal{O}_1(\Delta y; y)|_p / |\Delta y|_p = 0$$

for every $y \in L^p(T, E_n)$, uniformly in $\omega^{\square} \in \Omega$; hence $\mathcal{O}_1(\Delta y; y) = \tilde{F}_y(y, \omega^{\square})\Delta y$ $(\Delta y \in L^p(T, E_n))$, and $\tilde{F}_y(y,\omega^{\square})$ is the operator $\Delta y \to \int_T \tilde{f}_v(\cdot,\tau,y(\tau),\omega^{\square})\Delta y(\tau)\,d\tau$. Thus $\tilde{F}_y(y,\omega^{\square})$ and $\tilde{F}_y(\bar{y},0^{\square})$ have the form indicated in the statement of the Lemma.

The argument we have used to prove the existence of $\tilde{F}_y$ via inequality (7.1.2) and Assumption (4.1.3) can be used to show that

$|\tilde{F}_y(y_1, \omega^{\square}) - \tilde{F}_y(y_2, \omega^{\square})| \to 0$ as $y_2 \to y_1$ in $\mathcal{Y}$, uniformly in $\omega^{\square}$. Thus $y \to \tilde{F}(y,\omega^{\square})$ and $y \to \tilde{F}_y(y,\omega^{\square})$ are continuous at each y, uniformly in $\omega^{\square} \in \Omega$. Similar arguments show that $\omega^{\square} \to F(y,\omega^{\square})$ and $\omega^{\square} \to \tilde{F}_y(y,\omega^{\square})$

are continuous for each $y \in \mathcal{Y}$, whence we conclude that $(y,\omega^{\square}) \to \tilde{F}(y,\omega^{\square})$ and $(y,\omega^{\square}) \to \tilde{F}_y(y,\omega^{\square})$ exist and are continuous on $\mathcal{Y} \times \Omega$. Finally, the existence of $\tilde{F}_{\omega^{\square}}(y, 0^{\square})$ follows from that of $\tilde{f}_{\omega^{\square}}(t,\tau,y(\tau),0^{\square})$ and the bounds in (4.1.4). Thus $(y,\omega^{\square}) \to \tilde{F}(y,\omega^{\square})$ has a (total) derivative at $(\bar{y}, 0^{\square})$.

The same conclusions can be reached by similar arguments when $\mathcal{Y} = C(T, E_n)$. QED

Lemma 7.2 The mapping $I - F_y(\bar{y}, \bar{\sigma}, \bar{b})$ is a linear homeomorphism of $\mathcal{Y}$ onto $\mathcal{Y}$, and statement (4.2.1) is valid.

Proof: We have shown in Lemma 7.1 that

$(F_y(\bar{y}, \bar{\sigma}, \bar{b})\Delta y)(t) = \int k(t,\tau)\Delta y(\tau)d\tau \quad (t \in T, \Delta y \in \mathcal{Y})$.

By Assumption 4.1, k is measurable on $T \times T$ and $|k(t,\tau)|$ is bounded by $\tilde{\psi}(t,\tau) = (1 + |\bar{y}(\tau)|^{\alpha})\psi_1(t,\tau)$ for $\mathcal{Y} = L^p(T, E_n)$. We verify then, as in Lemma 7.1, that $\int |\tilde{\psi}(\cdot,\tau)|_p^{p/(p-1)} d\tau < \infty$. It follows [12, p.518] that $F_y(\bar{y}, \bar{\sigma}, \bar{b})$ is a compact operator on $L^p(T, E_n)$. Similarly, if $\mathcal{Y} = C(T, E_n)$, the family of functions $t \to \int k(t,\tau)\Delta y(\tau)d\tau$ corresponding to all Δy such that $\max_{t \in T} |\Delta y(t)| \leq 1$ is uniformly bounded and has the common modulus of continuity Φ. Thus, in both cases, $F_y(\bar{y},\bar{\sigma}, \bar{b})$ is a compact operator. It follows, therefore, from (4.1.5) that $I - F_y(\bar{y},\bar{\sigma},\bar{b})$ is a linear homeomorphism of $\mathcal{Y}$ onto $\mathcal{Y}$ [12, Theorem 5, p. 579].

Let $K = F_y(\bar{y},\bar{\sigma},\bar{b})$ and $K^* = (I - K)^{-1} - I$.

For $\mathcal{Y} = L^p(T, E_n)$, the arguments of [14, pp. 157 - 160] (applying to the case $n = 1$, $p = 2$) can be suitably generalized to prove that K^* is an integral operator such that $(K^* \Delta y)(t) = \int k^*(t,\tau)\Delta y(\tau)d\tau$ $(t \in T, \Delta y \in \mathcal{Y})$, where k^* is as

described in (4.2.1). (These arguments, in their generalized form, are based on approximating the function $\tau \to k(\cdot,\tau)$ in $L^{p/(p-1)}(T, L^p(T, E_n))$ by finite sums of the form $\Sigma\alpha_j(\tau)\beta_j(\cdot)$). Finally, since k^i_j $(i = 1,\ldots,m)$ are independent of t and $K^* = K + KK^*$, the k^{*i}_j $(i = 1,\ldots,m)$ are also independent of t.

For $\mathcal{Y} = C(T, E_n)$, we observe that since $K^* = K + KK^*$ and K is compact, so is K^*. There exist, therefore [15, Proposition 9.5.17, p. 665], a measurable $k^\# = (k^{\#i}_j)(i, j = 1,\ldots,n)$ on $T \times T$ and a nonnegative regular Borel measure μ on T such that

$$(K^* \Delta y)(t) = \int k^\#(t,\tau)\Delta y(\tau)\mu(d\tau) \qquad (t \in T,\ \Delta y \in \mathcal{Y})$$

and $\sup_{t\in T} \int_T |k^\#(t,\tau)|\mu(d\tau) < \infty$. Our conclusions about k^* will follow directly from the Radon-Nikodym theorem once we prove that, for all $t \in T$, the measure $A \to \int_A k^\#(t,\tau)\mu(d\tau)$ is absolutely continuous with respect to our original measure $A \to \int_A d\tau$. This we can do by observing that if $K\Delta y_i \xrightarrow[i\to\infty]{} 0$ in $\mathcal{Y}$, so does $K^*\Delta y_i = (I + K^*)K\Delta y_i$; and then considering any sequence $\{A_i\}$ of Borel sets in T such that $|A_i| \xrightarrow[i\to\infty]{} 0$ and "approximating" their characteristic functions with continuous functions a_i such that $0 \le a_i(t) \le 1$, $a_i(t) = 1$ on C_i, $a_i(t) = 0$ on $T - G_i$, where $C_i \subset A_i \subset G_i$, C_i are closed, G_i are open, and $\mu(G_i - C_i) + |G_i - C_i| \to 0$ as $i \to \infty$. QED

7.3 Completion of proof of Theorem 4.2. Lemmas 7.1 and 7.2 show that Theorem 2.3 is applicable to the control problem as defined in §4, and that statement (4.2.1) is valid. We have, for $q = (\sigma,b)$, $\sigma_{11} = \sigma$, $b_{11} = b$,

$$DF(\bar{y},\bar{q};\ q - \bar{q}) = DF(\bar{y},\bar{\sigma},\bar{b}\ ;\ (\sigma,b) - (\bar{\sigma},\bar{b})) = \tilde{F}_{\omega 11}(\bar{y},0)$$

and

$$c^i(y,q) = F^i(y,q) \qquad (i = 1,\ldots,m).$$

Let K^* be the linear operator on $\mathcal{Y}$ defined by k^* and let $\hat{\lambda} = (\lambda,0,\ldots,0) \in E_n$. Then $K^* = (I - F_y(\bar{y},\bar{q}))^{-1} - I$ and, applying Lemma 7.1, relation (2.3.2) can be rewritten as

$$\hat{\lambda}\cdot\{F_y(\bar{y},\bar{q})(I - F_y(\bar{y},\bar{q}))^{-1}DF(\bar{y},\bar{q};\ q - \bar{q}) + DF(\bar{y},\bar{q};\ q - \bar{q})\} \tag{7.3.1}$$

$$= \hat{\lambda}\cdot\ (I + K^*)DF(\bar{y},\bar{q};\ q)$$

$$= \sum_{j=1}^{m} \lambda^j \int_T \{f^j(\theta,\bar{y}(\theta),\sigma(\theta) - \bar{\sigma}(\theta),\bar{b}) + Df^j(\theta,\bar{y}(\theta),\bar{\sigma}(\theta),\bar{b};b-\bar{b})\}d\theta$$

$$+ \sum_{i=1}^{m}\sum_{j=1}^{n} \lambda^i \int_{T\times T} k^{*i}_j(\tau)\{f^j(\tau,\theta,\bar{y}(\theta),\sigma(\theta)-\bar{\sigma}(\theta),\bar{b})+Df^j(\tau,\theta,\bar{y}(\theta),\bar{\sigma}(\theta),\bar{b};b-\bar{b})\}d\tau d$$

$$= \int_T d\theta \int_T \zeta(\tau)\cdot\{f(\tau,\theta,y(\theta),\sigma(\theta)-\bar{\sigma}(\theta),\bar{b})+Df(\tau,\theta,\bar{y}(\theta),\bar{\sigma}(\theta),\bar{b};b-\bar{b})\}d\tau$$

$$= \int_T H_1(\sigma(\theta),\theta)d\theta - \int_T H_1(\bar{\sigma}(\theta),\theta)d\theta + H_2(b) \geq 0$$

for all $(\sigma,b) \in \mathcal{Y}^{\#} \times B$. In particular, for $\sigma = \bar{\sigma}$, $H_2(b) \geq 0 = H_2(\bar{b})$ for all $b \in B$.

It remains now to prove relation (1). Let $\mathcal{Y}^{\#}_{\infty} = \{\rho_1,\rho_2,\ldots\}$, $i \in \{1, 2,..\}$, E be an arbitrary measurable subset of T, $b = \bar{b}$, and $\sigma(t) = \rho_i(t)$ for $t \in E$, $\sigma(t) = \bar{\sigma}(t)$ for $t \in T - E$. Then relation (7.3.1) yields

$$\int_E \{H_1(\rho_i(\theta),\theta) - H_1(\bar{\sigma}(\theta),\theta)\}d\theta \geq 0,$$

where $H_1(r,\theta) = H_1(s_r,\theta)$ and s_r is a measure concentrated at r with mass 1. It follows that for each i there exists a subset T_i of T, of measure $|T|$, such that

$$H_1(\rho_i(\theta),\theta) \geq H_1(\bar{\sigma}(\theta),\theta) \quad \text{for all} \quad \theta \in T_i.$$

Then, for $T' = \bigcap_{i=1}^{\infty} T_i$,

$$(7.3.2) \quad H_1(r,\theta) = \int_T \zeta(\tau)\cdot g(\tau,\theta,\bar{y}(\theta),r,\bar{b})d\tau \geq H_1(\bar{\sigma}(\theta),\theta)$$

for all $\theta \in T'$ and $r \in R^*(\theta)$. We verify, using properties of k^* described in (4.2.1) and the bounds on g described in Assumption (4.1.4), that $\tau \to \zeta(\tau) \cdot g(\tau,\theta,\bar{y}(\theta),r,\bar{b})$ is bounded for all r and almost all θ by an integrable function of τ. Since, furthermore, it is also continuous in r, we conclude that relation (7.3.2) is valid for almost all θ and all $r \in \bar{R}^*(\theta)$ and, integrating both sides with respect to any $s \in S^*(\theta)$, that relation (1) is valid.

When $R^\#(t) = R$ on T, we may choose as $\mathcal{L}_\infty^\#$ any set of constant functions from T to R whose images form a dense subset of R; then $\bar{R}^*(t) = R$ and $S^*(t) = S$ for all $t \in T$. QED

8. Proof of Theorem 1.3.5. Those parts of Theorem 1.3.5 that refer to the existence and necessary conditions follow directly from Theorems 3.2 and 4.2 whose assumptions are weaker. The statement asserting the existence of "approximating" sequences $\{\rho_j\}$ and $\{y_j\}$ will follow from Theorem 3.3 if we can prove that the equation $y = F(y,\rho,\bar{b})$ admits at least one solution y in $L^1(T,E_n)$ for each $\rho \in \mathcal{R}^{\#}$. This last statement follows from the fixed point theorem; indeed, for each ρ, the mapping $y \to F(y,\rho,\bar{b})$ is continuous in $C(T,E_n)$ and, because of the boundedness and the uniform continuity of g, the image of this mapping is contained in a convex and compact set of functions in $C(T,E_n)$ with a common bound and a common modulus of continuity. QED

References

[1] A. Friedman — Optimal control for hereditary processes, Arch.Ration. Mech. Anal. 15(1964), pp.396-416.

[2] M.N. Oğuztöreli — Time-Lag Control Systems,Academic Press,New York,1966.

[3] A. Halanay — Optimal controls for systems with time-lag, S.I.A.M. J. Control 6(1968),pp.215-234.

[4] A.G. Butkovskii — The maximum principle for optimum systems with distributed parameters, Autom. and Remote Control 22(1961), pp.1156-1176.

[5] J. Warga — Functions of relaxed controls, S.I.A.M. J. Control, 5(1967), pp. 628 - 641.

[6] _____ — Restricted minima of functions of controls, S.I.A.M. J. Control 5(1967), pp. 642 - 656.

[7] _____ — Minimizing variational curves restricted to a preassigned set, Trans. Amer. Math. Soc. 112, (1964), pp. 432 - 455.

[8] _____ — Unilateral variational problems with several inequalities, Michigan Math. J. 12(1965), pp. 449 - 480.

[9] _____ — On a class of minimax problems in the calculus of variations, Michigan Math. J. 12(1965), pp. 289 - 311

[10] E.J. Mc Shane — Necessary conditions in generalized - curve problems of the calculus of variations, Duke Math. J., 7(1940), pp. 1 - 27.

[11] J. Dieudonné — Foundations of Modern Analysis, Academic Press, New York, 1960.

[12] N. Dunford and J. Schwartz — Linear Operators, Part I, Interscience, New York, 1964.

[13] M.A. Krasnoselskii and Ya. B. Rutickii — Convex functions and Orlicz spaces, P. Noordhoff Ltd., Groningen, 1961.

[14] F. Riesz and B. Sz. - Nagy — Functional Analysis, F. Ungar Publ. Co., New York, 1955.

[15] R.E. Edwards — Functional Analysis, Holt, Rinehart and Winston, New York, 1965.

Northeastern University
Boston, Massachusetts

ORIGINAL MINIMIZING CONTROLS FOR INTEGRAL EQUATIONS

By J. Warga
Professor, Northeastern University
Boston, Massachusetts

NASA Grant NGR 22-011-020

ORIGINAL MINIMIZING CONTROLS FOR INTEGRAL EQUATIONS*

by J. Warga

1. Introduction . We consider a class of variational problems defined by the Uryson-type integral equation

$$y(t) = (y^1(t),\dots,y^n(t)) = \int_T g(t,\tau,y(\tau),\rho(\tau), b)\,\mu(d\tau)\ (t\in T)$$

where ρ is chosen from a given set $\mathcal{R}$ of "original" (unrelaxed) controls and b from a given convex set B of control parameters. We have investigated, in [1] , a related problem in which the set $\mathcal{R}$ was imbedded in a set $\mathcal{S}$ of measurable relaxed controls, and have discussed the existence of a minimizing relaxed control, its approximation by original controls, and necessary conditions for a relaxed minimum. Since, as it is well known from the control theory of ordinary differentia equations, the existence of a minimizing original (unrelaxed) control cannot be assured, except under very restrictive conditions we begin the present study with the a priori assumption that there exist an original control $\bar\rho \in \mathcal{R}$ and a parameter $\bar b \in B$ that yield a minimizing solution of the variational problem in $\mathcal{R}\times B$. We then show, applying certain results of [2] , that the necessary conditions for minimum derived in [1] (generalizations of the Weierstrass E-condition and of the transversality conditions) remain essentially valid in the present context. Our present results are limited to the case where T is the closure of a bounded open set in the Euclidean ℓ-space E_ℓ and μ is absolutely continuous with respect to the Lebesgue measure (whereas in [1] T was only assumed to be metric and compact, with an

*This research was supported by N.A.S.A. Grant NGR 22-011-020.

appropriate measure); within this context, however, the present results generalize the necessary conditions of [1, Theorem 4.2] in that the given class of controls may, but need not, consist of relaxed controls and the remaining assumptions are also slightly weaker.

References to other related work can be found in [1].

2. Necessary conditions for minimum. Let T be the closure of a bounded open subset of E_ℓ, R a metric space, B a convex subset of a real linear space, B_1 a closed subset of E_m, $n \geq m$, $g = (g^1,\ldots,g^n)$, and $(t,\tau,v,r,b) \to g(t,\tau,v,r,b) : T \times T \times E_n \times R \times B \to E_n$. We assume that $g^i(t,\tau,v,r,b) = g^i(\tau,v,r,b)$ $(i = 1,\ldots,m)$ are independent of t.

For $\tilde{\rho}: T \to R$, $\rho_i: T \to R$ $(i = 1,\ldots,k)$ and disjoint subsets $A_1,\ldots,A_k$ of T, we define $\rho = [\rho_i, A_i (i = 1,\ldots,k); \tilde{\rho}] : T \to R$ by $\rho(t) = \rho_i(t)$ for $t \in A_i (i = 1,\ldots,k)$ and $\rho(t) = \tilde{\rho}(t)$ for $t \in T - \bigcup_{i=1}^{k} A_i$. Let $\mathcal{R}$ be any class of Lebesgue measurable mappings from T to R with the property that, for every set A that is a finite or denumerable union of intervals in E_ℓ, $(\rho_1 \in \mathcal{R}, \rho_2 \in \mathcal{R})$ implies $[\rho_1, A; \rho_2] \in \mathcal{R}$.

Let μ' be a Lebesgue integrable scalar function on T (viewed as a subset of E_ℓ), with $\mu'(\tau) > 0$ on T, and let μ be a positive measure defined on the class of Lebesgue measurable subsets of T by the relation $\mu(A) = \int \mu'(\tau) d\tau$, where $d\tau$ refers to the Lebesgue measure in E_ℓ. Let $\mathcal{Y}^A$ be either the Banach space $L^p(T, \mu; E_n)$ (of functions from T to E_n) for

$1 < p < \infty$ or the space $C(T, E_n)$ (of continuous functions from T to E_n), each with the usual norm. We consider the integral equation

$$y(t) = \int_T g(t,\tau,y(\tau),\rho(\tau),b)\,\mu(d\tau) \qquad (t \in T) \tag{2.0.1}$$

for $(\rho,b) \in \mathcal{R} \times B$. A point $(y,\rho,b) \in \mathcal{Y} \times \mathcal{R} \times B$ satisfying equation (2.0.1) is an "admissible" solution if $(y^1,y^2,\ldots y^m) \in B_1$ (observe that y^i $(i = 1,\ldots,m)$ are independent of t). An admissible solution $(\bar{y},\bar{\rho},\bar{b})$ is a "minimizing" solution if $\bar{y}^1 \leq y^1$ for every admissible (y,ρ,b).

Our purpose is to derive conditions satisfied by a minimizing solution $(\bar{y},\bar{\rho},\bar{b})$ that generalize the Weierstrass E-condition (the maximum principle) and the transversality conditions.

We shall use the term "measurable" in the sense of the Lebesgue measure on E_ℓ when referring to subsets of T or functions on T, and in the sense of the corresponding product measure with respect to $T \times T$. We represent by $|a|$ the norm of an element of a normed linear space. If $\mathcal{X}$ and $\mathcal{Z}$ are Banach spaces, $\Gamma \subset \mathcal{X}$ and $x \to h(x) : \Gamma \to \mathcal{Z}$, we define the derivative $h_x(x_1)$ as a linear operator from $\mathcal{X}$ to $\mathcal{Z}$ such that $|h(x) - h(x_1) - h_x(x_1)(x - x_1)| = o(|x - x_1|)$ for all $x \in \Gamma$. We denote by $h_{(x,y)}$, h_x, h_y the derivative and the partial derivatives, respectively, of a function $(x,y) \to h(x,y)$ from a subset of a Banach space to a Banach space. The symbol I represents the identity operator on $\mathcal{Y}$

$\mathcal{B}(\mathcal{X}, \mathcal{Z})$ the linear space of bounded linear operators from a Banach space $\mathcal{X}$ to a Banach space $\mathcal{Z}$ with the metric topology induced by the operator norm, A° the interior and $\bar{A}$ the closure of A, and ω^{α} an array $(\omega^{ij})(i, j = 1,\ldots,m)$. If $h: \mathcal{X} \times B \to \mathcal{Z}$, where $\mathcal{Z}$ is a Banach space, we write $Dh(x, b_1; b - b_1)$ for the one-sided derivative $\lim_{\alpha \to +0} \frac{1}{\alpha}(h(x, b_1 + \alpha(b - b_1)) - h(x, b_1))$.

Assumption 2.1. For every fixed choice of b^{α}, with elements $b^{ij} \in B$, the following conditions are satisfied:

(2.1.1) there exists $\theta_{max} \in (0, 1/m^2]$ such that, for $\mathcal{T}(b^{\alpha}) = \mathcal{T} = \{\theta^{\alpha} | 0 \leq \theta^{ij} \leq \theta_{max}\} \subset E_{m^2}$, the function $(t, \tau, v, r, \theta^{\alpha}) \to g^{\#}(t, \tau, v, r, \theta^{\alpha}) = g(t, \tau, v, r, \bar{b} + \sum_{i,j=1}^{m} \theta^{ij}(b^{ij} - \bar{b}))$: $T \times T \times E_n \times R \times \mathcal{T} \to E_n$ has a derivative with respect to (v, θ^{α}) everywhere, and $g^{\#}$, $g^{\#}_{v}$ and $g^{\#}_{\theta\sigma}$ are measurable in (t, τ) for every (v, r, θ^{α}) and continuous in (v, r, θ^{α}) for every (t, τ);

(2.1.2) (1) if $\mathcal{Y} = L^p(T, \mu; E_n)$ then there exist measurable positive ψ_o and ψ_1 on $T \times T$ and numbers α and β such that $0 \leq \alpha < p-1$, $0 \leq \beta < p$,

$$\int_T \{ \int_T |\psi_o(t , \tau)|^p \mu(dt)\}^{1/(p-\beta)} \mu(d\tau) < \infty ,$$

$$\int_T \{ \int_T |\psi_1(t , \tau)|^{p/(p - 1 - \alpha)} \mu(d\tau)\}^{(p - 1 - \alpha)} \mu(dt) < \infty,$$

$$\int_T \{ \int_T |\psi_1(t , \tau)|^p \mu(dt)\}^{1/(p - 1 - \alpha)} \mu(d\tau) < \infty ,$$ and, for

all $(t, \tau, v, r, \theta^{\alpha}) \in T \times T \times E_n \times R \times \mathcal{T}$,

$$|\hat{g}(t , \tau , v , r , \theta^{\alpha})| \leq (1 + |v|^{\beta})\psi_o(t , \tau) \quad \text{for}$$

$\hat{g} = g^{\#}$ and $g^{\#}_{\theta^{\alpha}}$

and

$$|g^{\#}_v(t , \tau , v , r , \theta^{\alpha})| \leq (1 + |v|^{\alpha})\psi_1(t , \tau) ;$$

(2) if $\mathcal{Y} = C(T, E_n)$ then there exist a compact set D in E_n containing $\{\bar{y}(t) | t \in T\}$ in its interior and an integrable scalar ψ on T such that, for $\hat{g} = g^{\#}$, $g^{\#}_v$, and $g^{\#}_{\theta^{\alpha}}$, we have

$$|\hat{g}(t, \tau , v , r , \theta^{\alpha})| \leq \psi(\tau)$$

for all $(t , \tau , v , r , \theta^{\alpha}) \in T \times T \times D \times R \times \mathcal{T}$. Furthermore, there exists a positive function $h \to \Phi(h)$ such that $\lim_{h \to +0} \Phi(h) = 0$ and, for $t_1 , t_2 \in T$ and $\hat{g} = g^{\#}$ and $g^{\#}_v$,

$$\int \operatorname*{Sup}_{D \times R \times \mathcal{T}} |\hat{g}(t_1 , \tau , v , r , \theta^{\alpha}) - \hat{g}(t_2 , \tau , v , r , \theta^{\alpha})| \mu(d$$

$$\leq \Phi(|t_1 - t_2|);$$

(2.1.3) for $k(t , \tau) = g_v(t , \tau , \bar{y}(\tau) , \bar{\rho}(\tau) , \bar{b})$ on $T \times T$, the integral equation

$$w(t) = \int_T k(t , \tau)\, w(\tau)\, \mu(d\tau) \qquad (t \in T)$$

has only the solution $w(\cdot) = 0$ in $\mathcal{Y}$.

2.2. Resolvent kernel. It follows from Assumption 2.1 (and can be proven exactly as in [1 , Lemma 7.2]) that there exists a measurable real matrix-valued function $\mathbf{k}^* = (k^{*i}_j)(i , j = 1,\ldots,n$ on $T \times T$ (a resolvent kernel of k) such that, for every $h \in \mathcal{Y}$ the relations

$$w(t) = \int_T k(t , \tau) w(\tau) \mu(d\tau) + h(t) \qquad (t \in T)$$

and

$$w(t) = \int_T k^*(t , \tau) h(\tau) \mu(d\tau) + h(t) \qquad (t \in T)$$

are equivalent in $\mathcal{Y}$, $k^{*i}_j(t , \tau) = k^{*i}_j(\tau)$ are independent of t for $i = 1,\ldots,m$, $\int_T \{ \int_T |k^*(t , \tau)|^p \mu(dt)\}^{1/(p-1)} \mu(d\tau) < \infty$ for $\mathcal{Y} = L^p(T , \mu ; E_n)$ and $\int_T \operatorname{Sup}_{t \in T} |k^*(t , \tau)| \mu(d\tau) < \infty$ for

$\mathcal{Y} = C(T, E_n)$.

We can now state our basic results.

Theorem 2.3. Let $(\bar{y}, \bar{\rho}, \bar{b})$ be a minimizing solution, let Assumption 2.1 be satisfied, and let k^* be a resolvent kernel of k. Let $\mathcal{R}_\infty$ be a denumerable subset of $\mathcal{R}$ containing $\bar{\rho}$, $R^*(t) = \{\rho(t) \mid \rho \in \mathcal{R}_\infty\}$ $(t \in T)$, K_1 a convex subset of some E_q, $\bar{\xi} \in K_1$, and $\phi: K_1 \to B_1$ a continuous mapping with a derivative at $\bar{\xi}$ and such that $\phi(\bar{\xi}) = (\bar{y}^1, \ldots, \bar{y}^m)$. Then

$$\text{either} \quad \phi'^1_\xi(\bar{\xi})\bar{\xi} = \operatorname*{Min}_{\xi \in K_1} \phi'^1_\xi(\bar{\xi})\xi ,$$

or there exist a nonvanishing $\lambda = (\lambda^1, \ldots, \lambda^m) \in E_m$ and $\gamma \geq 0$ such that, setting

$$\hat{\lambda} = (\lambda^1, \ldots, \lambda^m, 0, \ldots, 0) = (\lambda, 0, \ldots, 0) \in E_n ,$$

$$\zeta^j(\tau) = \sum_{i=1}^{m} \lambda^i k^{*i}_j(\tau) + \hat{\lambda}^j/\mu(T) \qquad (\tau \in T, \ j = 1, \ldots, n) ,$$

$$\zeta(\tau) = (\zeta^1(\tau), \ldots, \zeta^n(\tau)) \qquad (\tau \in T) ,$$

$$H_1(r, \tau) = \int_T \zeta(t) \cdot g(t, \tau, \bar{y}(\tau), r, \bar{b}) \mu(dt) \qquad (\tau \in T, \ r \in R$$

and

$$H_2(b) = \int_{T \times T} \zeta(t) \cdot Dg(t, \tau, \bar{y}(\tau), \bar{\rho}(\tau), \bar{b}; b - \bar{b}) \mu(dt) \mu(d\tau) \qquad (b \in B),$$

the following conditions are satisfied:

(The Weierstrass E-condition)

$$(1) \qquad H_1(\bar{\rho}(\tau)\ ,\ \tau) = \operatorname*{Min}_{r\ \varepsilon\ R^*(\tau)} H_1(r,\ \tau) \quad \text{for } \mu\text{-almost al}$$

$$\tau \quad \text{in} \quad T\ ,$$

(Transversality conditions)

$$(2) \qquad \operatorname*{Min}_{b\ \varepsilon\ B} H_2(b) = H_2(\bar{b}) = 0\ ,$$

and

$$(3) \qquad (\gamma\delta_1 - \lambda)\ \phi_\xi(\bar{\xi})\,\bar{\xi} = \operatorname*{Min}_{\xi\ \varepsilon\ K_1} (\gamma\delta_1 - \lambda)\cdot\phi_\xi(\bar{\xi})\,\xi\ ,$$

where $\delta_1 = (1\ ,\ 0,..,0) \varepsilon E_m$.

In particular, if R is separable and $\mathcal{R}$ contains all constant functions from T to R, we can replace $R^*(\tau)$ by R in relation (1).

3. Proof of Theorem 2.3.

3.1 Regular sequences, admissible controls, the sets T^* and $R^*(t^*)$. Let $|A| = \int_A d\tau$ represent the Lebesgue measure of $A \subset T$, diam(A) the diameter of A and $S(A\ ,\ \delta)$ the closed δ-neighborhood of A. A sequence $\{M_j\}_{j=1}^{\infty}$ of closed subsets

of T is "regular" at $\bar{t}$ (covers $\bar{t}$ in the sense of Vitali [3 , p.212]) if $\mathrm{diam}(M_j) \to 0$ as $j \to \infty$, $\bar{t} \in M_j$, and $|S(M_j , 3\,\mathrm{diam}(M_j))| \leq \alpha |M_j|$ $(j = 1,2,\ldots)$ for some $\alpha > 0$.

For any μ-integrable function $\tau \to f(\tau)$ from T into some Banach space, let $T'_\mu(f)$ be the set of all the points t^* in T such that $|f(t^*)| < \infty$ and

$$\alpha = \lim_{j\to\infty} \frac{1}{\mu(M_j)} \int_{M_j} f(\tau)\mu(d\tau) = f(t^*)$$

for all sequences $\{M_j\}$ that are regular at t^*. Since $\alpha = \lim_{j\to\infty} \frac{1}{|M_j|} \int_{M_j} f(\tau)\mu'(\tau)d\tau \Big/ \frac{1}{|M_j|} \int_{M_j} \mu'(\tau)d\tau$, it is well known (proof as in [3,Th. 8, p. 217]) that $|T'_\mu(f)| = |T|$; hence $\mu(T'_\mu(f)) = \mu(T)$. We write $T'_\mu(f_1 , f_2 , \ldots.)$ for $T'_\mu(f_1) \cap T'_\mu(f_2) \cap \ldots$.

If $\mathcal{Y} = L^p(T , \mu ; E_n)$ then it follows from Assumption 2.1 that, for all $\rho \in \mathcal{R}$, the function $\hat{g}(\rho)$ defined by

$$\hat{g}(\rho)(\tau) = g(\cdot, \tau , \bar{y}(\tau) , \rho(\tau) , \bar{b}) \qquad (\tau \in T)$$

is a μ-integrable function from T to $L^p(T , \mu ; E_n)$. We then set

$$T^* = \bigcap_{\rho \in \mathcal{R}_\infty} T'_\mu(\hat{g}(\rho) , \hat{g}(\bar{\rho})) \cap T^o . \tag{3.1.1}$$

If $\mathcal{Y} = C(T , E_n)$ then, for all $\rho \in \mathcal{R}$, the functions $g_1(\rho)$ and $g_2(\rho)$ from T to E_m, defined by

$$g_1^i(\rho)(\tau) = g^i(\tau , \bar{y}(\tau) , \rho(\tau), \bar{b}) \qquad (i = 1,..,m,\tau \in T)$$

and

$$g_2^i(\rho)(\tau) = \int_T \sum_{j=1}^{n} k^{*i}{}_j(t) g^j(t , \tau , \bar{y}(\tau) , \rho(\tau) , \bar{b}) \mu(dt)$$

$$(i = 1,..,m,\tau \in T)$$

are μ-integrable (since $t \to k^{*i}{}_j(t)$ are μ-integrable on T for $i = 1,...,m$ and $|g(t , \tau , \bar{y}(\tau) , \rho(\tau) , \bar{b})| \leq \psi(\tau)$ for all $t , \tau \in T$). We then set

$$(3.1.2) \qquad T^* = \bigcap_{\rho \in \mathcal{R}_\infty} T'_\mu(\hat{g}_1(\rho), \hat{g}_2(\rho), \hat{g}_1(\bar{\rho}), \hat{g}_2(\bar{\rho}), \psi) \cap T^o .$$

Thus, in both cases, $|T^*| = |T|$ and $\mu(T^*) = \mu(T)$.

We also set, for each $t^* \in T^*$,

$$R^*(t^*) = \{\rho(t^*) \mid \rho \in \mathcal{R}_\infty\} .$$

3.2 The collection $\mathcal{N}$ and the function G. Let

$$N_k = \bigcup_{0 \leq i \equiv k \pmod{m^2}} (2^{-i} , 2^{-i+1}] \quad (k = 1,2,..,m^2) ,$$

$\beta > 0$, $[0,\beta]^j = [0,\beta] \times ... \times [0,\beta]$ (j times),

$N_{k\beta} = (N_k \cap [0,\beta]) \times [0,\beta]^{\ell} - 1$, and $\gamma_k(t) = \sup_{\gamma > 0} \mu((t+N_{k\gamma}) \cap T^o)$ $(k=1,...,m^2,$

ORIGINAL MINIMIZING CONTROLS FOR INTEGRAL EQUATIONS

We define $N_k(t\ ,\ \alpha)$ $(t \in T^o\ ,\ k = 1,\ldots,m^2\ ,\ \alpha > 0)$ as $(t + N_{k\beta}) \bigcap T^o$, where β is chosen so that $\mu(N_k(t,\alpha)) = \mathrm{Min}(\alpha, \gamma_k(t))$.

We set $\mathcal{N}_k = \{N_k(t\ ,\ \alpha) \mid t \in T^*\ ,\ \alpha > 0\}$ $(k = 1,2,..,m^2)$ and $\mathcal{N} = \{\mathcal{N}_k \mid k = 1,2,\ldots,m^2\}$.

We can easily verify that, whenever $\{\alpha_j\}_{j=1}^{\infty}$ is a sequence decreasing to 0, the sequence $\{\bar{N}_k(t\ ,\ \alpha_j)\}_{j=1}^{\infty}$ is regular at t for every $t \in T^* \subset T^o$ and $k=1,2,..,m^2$.

For any fixed choice of $t^{ij} \in T^*$, $\rho^{ij} \in \mathcal{R}_\infty$, and $b^{ij} \in B$ $(i,j = 1,..,m)$, let $\mathcal{T} = \mathcal{T}(b^\omega)$ (as in Assumption 2.1), and let

$$\Omega = \Omega(t^\omega) = \{\omega^\omega \mid \omega^{ij} \geq 0 \ \ (i,j = 1,..,m) \ \text{ and the sets}$$

$$N_{mj - m + i}(t^{ij}\ ,\ \omega^{ij}) \ \text{ are disjoint}\} .$$

We set $\rho'(\omega^\omega) = [\rho^{ij}\ ,\ N_{mj - m + i}(t^{ij}\ ,\ \omega^{ij})(i,j = 1,\ldots,m);\ \bar{\rho}]$ and define $g^\#$ as in Assumption 2.1. Finally, we verify as in [1, proof of Lemma 7.1] that there exists a neighborhood $\Gamma = \Gamma_y \times \Gamma_\omega \times \Gamma_\theta$ of $(\bar{y}\ ,\ 0^\omega, 0^\omega)$ in $\mathcal{Y} \times \Omega \times \mathcal{T}$ such that the relation

$$G(y,\omega^\omega,\theta^\omega)(t) = \int_T g^\#(t,\tau,y(\tau),\rho'(\omega^\omega)(\tau),\theta^\omega)\mu(d\tau) \quad (t \in T)$$

defines a mapping $G: \Gamma \to \mathcal{Y}$.

Lemma 3.3. The functions $(y,\omega^\omega,\theta^\omega) \to G(y,\omega^\omega,\theta^\omega)$:

$$\Gamma \to \mathcal{Y} \ \text{ and } \ (y,\omega^\omega,\theta^\omega) \to G_y(y,\omega^\omega,\theta^\omega): \ \Gamma \to \mathcal{B}(\mathcal{Y},\mathcal{Y})$$

exist and are continuous. We have, for all $(y,\omega^{\circ},\theta^{\circ}) \in \Gamma$ and $\Delta y \in \mathcal{Y}$,

$$(3.3.1) \qquad (G_y(y,\omega^{\circ},\theta^{\circ})\Delta y)(t) = \int_T g_v^{\#}(t,\tau,y(\tau),\rho'(\omega^{\circ})(\tau),\theta^{\circ})\Delta y(\tau)\mu(d\tau)$$

$$(t \in T).$$

Proof. The arguments of [1 , Lemma 7.1] show that, for fixed ω°, the function $(y,\theta^{\circ}) \to G_y(y,\omega^{\circ},\theta^{\circ})\colon \Gamma_y \times \Gamma_\theta \to \mathcal{B}(\mathcal{Y},\mathcal{Y})$ exists, is continuous, and satisfies relation (3.3.1). Now let ω_1° and ω_2° be in Γ_ω and set $M_{1,2} = \{t \in T \mid \rho'(\omega_1^{\circ})(t) \neq \rho'(\omega_2^{\circ})(t)\}$. Then $\mu(M_{1,2}) \leq \sum_{i,j=1}^{m} |\omega_1^{ij} - \omega_2^{ij}|$ and, by (3.3.1),

$$((G_y(y,\omega_1^{\circ},\theta^{\circ}) - G_y(y,\omega_2^{\circ},\theta^{\circ}))\Delta y)(t) =$$

$$= \int_{M_{i,2}} (g_v^{\#}(t,\tau,y(\tau),\rho'(\omega_1^{\circ})(\tau),\theta^{\circ}) - g_v^{\#}(t,\tau,y(\tau),\rho'(\omega_2^{\circ})(\tau),\theta^{\circ}))\cdot\Delta y(\tau)\mu(d\tau)$$

$$((y,\omega_i^{\circ},\theta^{\circ}) \in \Gamma \ , \ \Delta y \in \mathcal{Y} \ , \ t \in T)\cdot$$

For $\mathcal{Y} = L^p(T,\mu;E_n)$, we have, therefore, in view of Assumption (2.1.2),

$$A_{1,2} = |(G_y(y,\omega_1^{\circ} \ , \ \theta^{\circ}) - G_y(y,\omega_2^{\circ},\theta^{\circ}))\Delta y|_p \leq \int_{M_{1,2}} 2\,|\psi_1(\cdot,\tau)|_p\,|\Delta y(\tau)|\mu(d\tau)$$

where the function $\tau \to |\psi_1(\cdot,\tau)|_p = \{\int_T |\psi_1(t,\tau)|^p \mu(dt)\}^{1/p}$ belongs to $L^{p/(p-1)}(T,\mu)$; hence

$$A_{1,2} \leq 2 \{ \int_{M_{1,2}} |\psi_1(\cdot,\tau)|_p^{p/(p-1)} \mu(d\tau) \}^{1-1/p} \cdot |\Delta y|_p .$$

Since $\mu(M_{1,2}) \to 0$ as $\omega_2' \to \omega_1^*$, it follows that $A_{1,2}/|\Delta y|_p$ also converges to 0, uniformly in (y,θ^*). Thus, when $\mathcal{Y} = L^p(T,\mu;E_n)$, the function $\omega^* \to G_y(y,\omega^*,\theta^*) : \Gamma_\omega \to \mathcal{B}(\mathcal{Y},\mathcal{Y})$ is continuous, uniformly in (y,θ^*), and we conclude that the function $(y,\omega^*,\theta^*) \to G_y(y,\omega^*,\theta^*)$ is continuous in $\mathcal{Y} \times \Omega \times \mathcal{T}$.

For $\mathcal{Y} = C(T, E_n)$, we have

$$|((G_y(y,\omega_1',\theta^*) - G_y(y,\omega_2',\theta^*)))\Delta y)(t)| \leq \int_{M_{1,2}} \psi(\tau)\mu(d\tau) \cdot |\Delta y|_\infty ,$$

and the argument can be continued as in the previous case. QED

Lemma 3.4. Let $\mathcal{Y} = L^p(T,\mu;E_n)$. For fixed $y \in \Gamma_y$, the function $(\omega^*,\theta^*) \to G_{\theta^*}(y,\omega^*,\theta^*) : \Gamma_\omega \times \Gamma_\theta \to \mathcal{B}(E_{m2},\mathcal{Y})$ exists and is continuous, and we have

$$(3.4.1) \qquad G_{\theta^\alpha}(y,\omega^\alpha,\theta^{\alpha'}) = \int_T g^{\#}_{\theta^\alpha}(\cdot,\tau,y(\tau),\rho'(\omega^\alpha),\theta^{\alpha'})\mu(d\tau) \quad (t \in T).$$

Proof. The existence of $G_{\theta^\alpha}(y,\omega^\alpha,\theta^{\alpha'})$, relation (3.4.1), and the continuity of $\theta^{\alpha'} \to G_{\theta^\alpha}(y,\omega^\alpha,\theta^{\alpha'})$ follow from Assumption (2.1) and, in particular, the continuity of $\theta^{\alpha'} \to g^{\#}_{\theta^\alpha}$ and the bounds. The continuity of $\omega^\alpha \to G_{\theta^\alpha}(y,\omega^\alpha,\theta^{\alpha'})$, uniformly in $(y,\theta^{\alpha'})$,

can be shown by the same arguments that were applied in Lemma 3.3, with $G_{\theta^\bullet}$ replacing G_y. QED

Lemma 3.5. Let $\mathcal{Y} = L^p(T,\mu;E_n)$ and let $0^\bullet$ be the origin of E_{m2}. Then the function $(y,\omega^\bullet,\theta^\bullet) \to G(y,\omega^\bullet,\theta^\bullet): \Gamma \to \mathcal{Y}$ has a (total) derivative at $(\bar{y},0^\bullet,0^\bullet)$ (relative to Γ), and

$$(3.5.1) \qquad G_{\omega ij}(\bar{y},0^\bullet,0^\bullet) = g(\cdot,t^{ij},\bar{y}(t^{ij}),\rho^{ij}(t^{ij}),\bar{b}) - g(\cdot,t^{ij},\bar{y}(t^{ij}),\bar{\rho}(t^{ij}),\bar{b}) \quad \text{in } \mathcal{Y}.$$

Proof. If follows from the bounds in Assumption (2.1.2) that $\tau \to |g^\#(\cdot,\tau,y(\tau),\rho'(\omega^\bullet)(\tau),\theta^\bullet)|_p$ is μ-integrable for all $(y,\omega^\bullet,\theta^\bullet) \in \Gamma$. Thus

$$G(\bar{y},\omega^\bullet,\theta^\bullet) = \int_T g^\#(\cdot,\tau,\bar{y}(\tau),\rho'(\omega^\bullet)(\tau),\theta^\bullet)\mu(d\tau) \quad \text{in } \mathcal{Y}$$

and

$$G(\bar{y},\omega^\bullet,0^\bullet) - G(\bar{y},0^\bullet,0^\bullet) = \sum_{i,j=1}^{m} \int_{M_{ij}} (g(\cdot,\tau,\bar{y}(\tau),\rho^{ij}(\tau),\bar{b}) - g(\cdot,\tau,\bar{y}(\tau),\bar{\rho}(\tau),\bar{b}))\mu(d\tau)$$

for $M_{ij} = M_{ij}(\omega^{ij}) = N_{mj-m+i}(t^{ij},\omega^{ij})$ $(i,j=1,..,m)$. Since, for each fixed i and j, the sequence $\{\bar{M}_{ij}(\alpha_k)\}_{k=1}^{\infty}$ is regular at t^{ij} if $\alpha_k \to +0$, the M_{ij} are disjoint, $t^{ij} \in T^*$ and $\mu(\bar{M}_{ij}) = \mu(M_{ij}) = \omega^{ij}$ for sufficiently small ω^{ij} $(i,j=1,..,m)$,

relation (3.5.1) follows directly and we also conclude that $\omega^\bullet \to G(\bar{y},\omega^\bullet,0^\bullet)$ has a derivative at $0^\bullet$. Since, by Lemma 3.4, $(\omega^\bullet,\theta^\bullet) \to G_{\theta^\bullet}(\bar{y},\omega^\bullet,\theta^\bullet): \Gamma_\omega \times \Gamma_\theta \to \mathcal{B}(E_{m2},\mathcal{Y})$ is continuous and there exists a convex neighborhood of $(0^\bullet,0^\bullet)$ relative to $\Gamma_\omega \times \Gamma_\theta$, we conclude that the function $(\omega^\blacksquare,\theta^\blacksquare) \to G(\bar{y},\omega^\square,\theta^\blacksquare)$ has a derivative at $(0^\blacksquare,0^\blacksquare)$. Finally, by Lemma 3.3, $(y,\omega^\blacksquare,\theta^\blacksquare) \to G_y(y,\omega^\blacksquare,\theta^\blacksquare): \Gamma \to \mathcal{B}(\mathcal{Y},\mathcal{Y})$ is continuous; hence $(y,\omega^\square,\theta^\square) \to G(y,\omega^\square,\theta^\square)$ has a derivative at $(\bar{y},0^\square,0^\square)$. QED.

3.6 Proof of Theorem 2.3 for $\mathcal{Y} = L^p(T,E_n)$ $(1<p<\infty)$.

By Lemmas 3.3 and 3.5, the function $(y,\omega^\bullet,\theta^\bullet) \to G(y,\omega^\blacksquare,\theta^\blacksquare): \Gamma \to \mathcal{Y}$ has a continuous partial derivative with respect to y and a (total) derivative at $(\bar{y},0^\bullet,0^\bullet)$. Furthermore, Assumption (2.1) implies (see [1,Lemma 7.2] for details) that $I-G_y(\bar{y},0^\blacksquare,0^\bullet)$ is a linear homeomorphism of $\mathcal{Y}$ onto $\mathcal{Y}$ and that

$$(3.6.1) \qquad (I-G_y(\bar{y},0^\bullet,0^\bullet))^{-1} = I + K^* ,$$

where

$$(3.6.2) \qquad (K^*\Delta y)(t) = \int_T k^*(t,\tau)\Delta y(\tau)\mu(d\tau) \quad \text{for all } t \in T \text{ and } \Delta y \in \mathcal{Y}.$$

It follows then from a variant of the implicit function theorem (using essentially the same arguments as in [4,p. 265]) and from

the representation of G_y, G_{θ^σ} and G_{ω^σ} in Lemmas 3.3, 3.4, and 3.5 that the equation

$$y = G(y, \omega^\sigma, \theta^\sigma)$$

has a unique solution $\eta(\omega^\sigma, \theta^\sigma) = (\eta^1(\omega^\sigma, \theta^\sigma), \ldots, \eta^n(\omega^\sigma, \theta^\sigma))$ in $\mathcal{Y}$ for all $(\omega^\sigma, \theta^\sigma)$ in some neighborhood Δ of $(0^\sigma, 0^\sigma)$ in $\Gamma_\omega \times \Gamma_\theta$ such that the function $(\omega^\sigma, \theta^\sigma) \to \eta(\omega^\sigma, \theta^\sigma)$ is continuous in Δ and has a derivative at $(0^\sigma, 0^\sigma)$, and

$$\eta_{\omega ij}(0^\sigma, 0^\sigma) = (I - G_y(\bar{y}, 0^\sigma, 0^\sigma))^{-1}(g(\cdot, t^{ij}, \bar{y}(t^{ij}), \rho^{ij}(t^{ij}), \bar{b}) - g(\cdot, t^{ij}, \bar{y}(t^{ij}), \bar{\rho}(t^{ij}), \bar{b})), \tag{3.6.3}$$

$$\eta_{\theta ij}(0^\sigma, 0^\sigma) = (I - G_y(\bar{y}, 0^\sigma, 0^\sigma))^{-1} \int_T g^{\#}_{\theta ij}(\cdot, \tau, \bar{y}(\tau), \bar{\rho}(\tau), \rho^\sigma) \mu(d \tag{3.6.4}$$

for $i, j = 1, .., n$.

Now let a function $(\rho, b) \to x(\rho, b): \mathcal{R} \times \mathcal{B} \to E_m$ be defined as follows:

if the equation

$$y(t) = \int_T g(t, \tau, y(\tau), \rho(\tau), b) \mu(d\tau) \qquad (t \in T)$$

has a unique solution $y(\cdot)$ in $\mathcal{Y}$, we set

$$x(\rho,b) = (y^1,\ldots,y^m)$$

(remembering that y^i $(i=1,..,m)$ are then independent of t); otherwise we set

$$x(\rho,b) = (\bar{y}^1 + 1,0,\ldots,0).$$

We also set, for all $t^* \in T^*$, $\mathcal{R}^*(t^*) = \mathcal{R}_\infty$.

We observe that, in particular, $x^i(\rho'(\omega^\sigma),\bar{b} + \sum_{i,j=1}^{m}\theta^{ij}(b^{ij}-\bar{b}) = \eta^i(\omega^\sigma,\theta^\sigma)$ in Δ $(i=1,\ldots,m)$.

We can now verify that $(\bar{\rho},\bar{b})$ yields the minimum of $x^1(\rho,b)$ on $\{(\rho,b) \in \mathcal{R} \times B | x(\rho,b) \in B_1\}$, that $(T^*,\mathcal{R}^*,\mathcal{N})$ define "local variations for x in $\mathcal{R} \times B$ at $(\bar{\rho},\bar{b})$" according to [2, Definition 2.1, p. 644], and that Theorem 2.2 of [2, p. 644] is therefore applicable. Furthermore, defining $Dx(\bar{\rho},\bar{b};t^*,r)$ as in [2,p. 643], we have

$$(3.6.5) \qquad Dx^i(\bar{\rho},\bar{b};t^*,r) = \eta^i_{\omega 1,1}(0^\sigma,0^\sigma) \qquad (i=1,..,m,t^* \in T^*,\ r \in R^*$$

where η is defined by choosing $t^{1,1} = t^*$, $\rho^{1,1}$ such that $\rho^{1,1}(t^*) = r$ and the other t^{ij},ρ^{ij} and b^σ arbitrarily; and

$$(3.6.6) \quad Dx^i(\bar{\rho},\bar{b};b-\bar{b}) = \eta^i_{\theta 1,1}(0^\sigma,0^\sigma) \qquad (i=1,..,m,b \in B),$$

where η is defined by choosing $b^{1,1} = b$ and t^σ,ρ^σ and the other b^{ij} arbitrarily. (The symbol $Dx^i(\bar{\rho},\bar{b};b)$ in the notation of [2] corresponds to $Dx^i(\bar{\rho},\bar{b};b-\bar{b})$ in our present notation).

It follows, by [2,Th. 2.2, p. 644], that either the first alternative of Theorem 2.3 is valid or there exist a nonvanishing λ in E_m and $\gamma \geq 0$ such that condition (3) of Theorem 2.3 is satisfied,

$$(3.6.7) \qquad \lambda \cdot Dx(\bar{\rho},\bar{b};t^*,r) \geq 0 \quad \text{for all} \quad t^* \in T^* \quad \text{and} \quad r \in R^*(t^*),$$

and

$$(3.6.8) \qquad \lambda \cdot Dx(\bar{\rho},\bar{b};b-\bar{b}) \geq 0 \quad \text{for all} \quad b \in B .$$

Relation (1) of Theorem 2.3 now follows from (3.6.7), taking account of (3.6.1),(3.6.2),(3.6.3), and (3.6.5). Similarly, relation (2) follows from (3.6.8), in view of relations (3.6.1), (3.6.2), (3.6.4), and (3.6.6).

It now remains to verify the statement that $R^*(t^*)$ can be replaced by R if R is separable and $\mathcal{R}$ contains all constant functions from T to R. In that case we can choose as $\mathcal{R}_\infty$ any denumerable subset of $\mathcal{R}$ containing $\bar{\rho}$ and a set of constant functions from T to R whose images form a dense subset of R. Then $\bar{R}^*(\tau) = R$ for all $\tau \in T^*$ and, since $r \to H_1(r,\tau)$ is continuous for all $\tau \in T$, we conclude that $\underset{r \in R^*(\tau)}{\text{Min}} H_1(r,\tau) = \underset{r \in R}{\text{Min}} H_1(r,\tau)$. Q

3.7. Proof of Theorem 2.3 for $\mathcal{Y} = C(T,E_n)$. The proof of Theorem 2.3 for $\mathcal{Y} = L^p(T,\mu;E_n)$ partly relied on the observation that $G(y,\omega^o,\theta^o)$ is the μ-integral over T of the function

$\tau \to g^{\#}(\cdot,\tau,y(\tau),\rho(\tau),\theta^{\circ}): T \to \mathcal{Y}$. For $\mathcal{Y} = C(T,E_n)$ this need not be (and in most cases of interest is not) the case since $g^{\#}(\cdot,\tau,y(\tau),\rho(\tau),\theta^{\circ})$ is not assumed continuous for fixed τ and θ°. We can circumvert this difficulty, however, since we are primarily interested in the first m components of y, and $(y^1,..,y^m)$ is constant for every solution y of $y = G(y,\omega^{\circ},\theta^{\circ})$. This remark motivates the ensuing arguments.

By Lemma 3.3, the functions G and G_y exist and are continuous on Γ. It follows, therefore, by the implicit function theorem, that the equation

$$y = G(y,\omega^{\circ},\theta^{\circ})$$

has a unique solution $\eta(\omega^{\circ},\theta^{\circ})$ in $\mathcal{Y}$ for $(\omega^{\circ},\theta^{\circ})$ in some neighborhood Δ of $(0^{\circ},0^{\circ})$ in $\Gamma_\omega \times \Gamma_\theta$ and that $(\omega^{\circ},\theta^{\circ}) \to \eta(\omega^{\circ},\theta^{\circ})$ $\Delta \to \mathcal{Y}$ is continuous.

Let $K = G_y(\bar{y},0^{\circ},0^{\circ})$, $K^* = (I-K)^{-1}-I$, and let P_m be the projection operator $(a^1,\ldots,a^n) = a \to P_m \cdot a = (a^1,..,a^m)$: $E_n \to E_m$.

Lemma 3.7.1. The function $(\omega^{\circ},\theta^{\circ}) \to P_m \cdot (I+K^*) \cdot G(\bar{y},\omega^{\circ},\theta^{\circ})$: $\Delta \to E_m$ has a derivative $\mathcal{B} = (\mathcal{B}^1,..\mathcal{B}^m)$ at $(0^{\circ},0^{\circ})$, and

$$(3.7.1.1)\quad \mathcal{B}^i \cdot (\omega^{\circ},\theta^{\circ}) = (\int_T g^{\#i}_{\theta^{\circ}}(\tau,\bar{y}(\tau),\bar{\rho}(\tau),0^{\circ})\mu(d\tau) + \int_{T\times T} k^{*i}(t) \cdot$$

$$g^{\#}_{\theta^{\circ}}(t,\tau,\bar{y}(\tau),\bar{\rho}(\tau),0^{\circ})\mu(dt)\mu(d\tau)) \cdot \theta^{\circ} +$$

$$\sum_{k,\ell=1}^{m}(g^i(t^{k\ell},\bar{y}(t^{k\ell}),\bar{\rho}(t^{k\ell}),\bar{b}) + \int_T k^{*i}(t)\cdot$$

$$g(t,t^{k\ell},\bar{y}(t^{k\ell}),\bar{\rho}(t^{k\ell}),\bar{b})\mu(dt))\omega^{k\ell} \quad (i=1,\ldots,m,\omega'\varepsilon E_{m2},\theta'\varepsilon E_{m2}),$$

where $k^{*i} = (k^{*i}_1,\ldots,k^{*i}_n)$.

Proof. We set $H(\omega',\theta') = P_m\cdot(I+K^*)\cdot G(\bar{y},\omega',\theta')$ and verify that

$$H^i(\omega',\theta') = \int_T g^{\#i}(\tau,\bar{y}(\tau),\rho'(\omega')(\tau),\theta')\mu(d\tau) +$$

$$+ \int_{T\times T} k^{*i}(t)\cdot g^{\#}(t,\tau,\bar{y}(\tau),\rho'(\omega')(\tau),\theta')\mu(dt)\mu(d\tau)$$

$$(i = 1,..,m).$$

Assumptions (2.1.1) and (2.1.2) imply that $(\omega',\theta') \to H_{\theta'}(\omega',\theta')$: $\Delta \to \mathcal{B}(E_{m2},E_m)$ exists, is continuous, and

$$H_{\theta'}(0',0')\cdot\theta' = \mathcal{D}\cdot(0',\theta'),$$

where $\mathcal{D}$ is defined as in (3.7.1.1).

We also observe that

$$H^i(\omega'',0'') - H^i(0'',0'') = \sum_{k,\ell=1}^{m}\int_{M_{k\ell}}(g^i(\tau,\bar{y}(\tau),\rho^{k\ell}(\tau),\bar{b})-g^i(\tau,\bar{y}(\tau),\bar{\rho}(\tau),\bar{b}))\mu(d$$

$$+ \sum_{k,\ell=1}^{m}\int_{M_{k\ell}}\mu(d\tau)\int_T k^{*i}(t)\cdot(g(t,\tau,\bar{y}(\tau),\rho^{k\ell}(\tau),\bar{b})-g(t,\tau,\bar{y}(\tau),\bar{\rho}(\tau),\bar{b}))\mu(dt),$$

where $M_{k\ell} = N_{m\ell-m+k}(t^{k\ell},\omega^{k\ell})$. Since $t^{k\ell} \in T^*$ for $k,\ell=1,..,m$, it follows from the definition of T^* that $H_{\omega^\circ}(0^\circ,0^\circ)$ exists and

$$H_{\omega^\circ}(0^\circ,0^\circ)\cdot\omega^\circ = \mathcal{O}\cdot(\omega^\circ,0^\circ).$$

We conclude that $H_{(\omega^\circ,\theta^\circ)}(0^\circ,0^\circ) = \mathcal{O}$. QED.

Lemma 3.7.2. There exists a constant c such that

$$\sup_{t\in T}|G(\bar{y},\omega^\circ,\theta^\circ)(t) - \bar{y}(t)| \leq c(|\omega^\circ|+|\theta^\circ|)$$

for all $(\omega^\circ,\theta^\circ)$ sufficiently close to $(0^\circ,0^\circ)$.

Proof. We have $\bar{y}(t) = G(\bar{y},0^\circ,0^\circ)(t)\ (t\in T)$ and, for all $t \in T$,

$$|G(\bar{y},\omega^\circ,\theta^\circ)(t) - \bar{y}(t)| \leq \int_T (g^\#(t,\tau,\bar{y}(\tau),\rho'(\omega^\circ)(\tau),\theta^\circ) - g^\#(t,\tau,\bar{y}(\tau),\rho'(\omega^\circ),0^\circ)\,\mu(d\tau)|$$

$$+\Big|\sum_{k,\ell=1}^{m}\int_{M_{k\ell}} (g^\#(t,\tau,\bar{y}(\tau),\rho^{k\ell}(\tau),0^\circ) - g^\#(t,\tau,\bar{y}(\tau),\bar{\rho}(\tau),0^\circ))\mu(d\tau)\Big|$$

$$= a + b\ ,$$

where $M_{k\ell} = N_{m\ell-m+k}(t^{k\ell},\omega^{k\ell})$. We observe that

$$a \leq \int_T \psi(\tau)\mu(d\tau)\cdot|\theta^\circ|$$

because $|g^\#_{\theta^\circ}(t,\tau,\bar{y}(\tau),r,\theta^\circ)| \leq \psi(\tau)$ everywhere, and that

$$b \leq 2 \sum_{k,\ell=1} \int_{M_{k\ell}} \psi(\tau)\mu(d\tau)$$

because $|g^{\#}(t,\tau,\bar{y}(\tau),r,b)| \leq \psi(\tau)$ everywhere. The conclusion of the lemma now follows directly, remembering that

$$\lim_{\omega^{k\ell}\to 0} \frac{1}{\omega^{k\ell}} \int_{M_{k\ell}} \psi(\tau)\mu(d\tau) = \psi(t^{k\ell}) \quad (k,\ell=1,\ldots,m)$$

since $t^{k\ell} \in T^* \subset T'_\mu(\psi)$ $\quad (k,\ell=1,..,m)$. QED

3.7.3 Completion of the proof. We now observe that, for $w = (\omega'',\theta'')$, $0' = (0'',0'')$, and for all $w \in \Delta$,

$$\eta(w) - G(\bar{y},w) = G(\eta(w), w) - G(\bar{y},w) = G_y(\tilde{\eta}(w),w)\cdot(\eta(w) - \bar{y});$$

hence

$$\eta(w) - \bar{y} = (I - G_y(\tilde{\eta}(w),w))^{-1}(G(\bar{y},w) - G(\bar{y},0')), \text{ where}$$

$\tilde{\eta}(w) \in [\bar{y}, \eta(w)] \subset \mathcal{Y}$. Thus

$$(3.7.3.1) \quad P_m\cdot(\eta(w) - \eta(0') - (I + K^*)\cdot(G(\bar{y},w) - G(\bar{y},0')))$$

$$= P_m\cdot(((I-G_y(\tilde{\eta}(w),w))^{-1} - (I-G_y(\bar{y},0'))^{-1})(G(\bar{y},w)$$

$$- G(\bar{y},0'))).$$

Since $w \to \eta(w)$ is continuous, and so is, by Lemma 3.3, $(y,w) \to G_y(y,w)$, it follows from Lemma 3.7.2 that the right hand side of (3.7.3.1) is $o(|w|)$. Thus, in view of Lemma 3.7.1,

the function $w \to P_m \cdot \eta(w)$ has a derivative at $0'$ and

$$(\eta^1, \ldots, \eta^m)_{(\omega^\circ, \theta^\circ)}(0^\circ, 0^\circ) = (P_m \cdot \eta)_w(0') = P_m \cdot (I + K^*) \cdot$$

$$\cdot G_{(\omega^\circ, \theta^\circ)}(\bar{y}, 0^\circ, 0^\circ) = \mathcal{D},$$

where $\mathcal{D}$ is defined by relation (3.7.1.1).

We can now complete the proof of the theorem exactly as we did in 3.6 for $\mathcal{Y} = L^p(T, \mu; E_n)$. QED

REFERENCES

[1] J. Warga — Relaxed Controls for Functional Equations, to appear in J. Funct. Anal.

[2] ———— Restricted Minima of Functions of Control, S.I.A.M. J. Control 5(1967), pp. 642-656.

[3] N. Dunford & J. Schwartz — Linear Operators, Part I, Interscience, New York, 1964.

[4] J. Dieudonné — Foundations of Modern Analysis, Academic Press, New York, 1960.

Northeastern University,
Boston, Massachusetts

APPLICATIONS OF HAMILTON-JACOBI THEORY TO PLANAR TRAJECTORY OPTIMIZATION

By S. K. Lakhanpal
Vanderbilt University
Nashville, Tennessee

NASA Research Grant NGR-43-002-015

APPLICATIONS OF HAMILTON-JACOBI THEORY TO PLANAR TRAJECTORY OPTIMIZATION*

By S. K. Lakhanpal
Vanderbilt University
Nashville, Tennessee

SUMMARY

The purpose of this paper is to study the application of Hamilton-Jacobi perturbation methods to the determination of the minimum fuel trajectory of a rocket moving in a plane under a central gravitational force and the thrust of an engine. First, a brief survey is given of the needed theorems from the calculus of variations and Hamilton-Jacobi theory. The problem is then formulated analytically and the multiplier rule and Weierstrass condition applied. The Hamiltonian is separated into base and perturbation parts. Two methods are given for obtaining a complete integral of the Hamilton-Jacobi partial differential equation for the base Hamiltonian. Jacobi's Theorem is applied to give a system of canonic constants for the base problem. The procedure for using these constants as canonic variables in the perturbing Hamiltonian is then developed.

INTRODUCTION

Many trajectory optimization problems are of the Mayer type in the calculus of variations, the classical theory being easily extended to include control variables. (See, for example, Hestenes, Ref. [1 or 2]). With differential constraints in normal form, the multiplier rule gives equations of extremals as canonical equations of a generalized Hamiltonian. Jacobi's theorem then gives a method of solution based on finding a complete integral of a partial differential equation. This theory is summarized briefly, without proofs, in the first part of this paper.

Low thrust rocket trajectory problems are analogous to perturbation problems of planetary theory, the thrust of the engine being considered as the perturbing force. William E. Miner [3] has developed this method extensively for three dimensional trajectories. The object here is to consider the simpler planar case and to study alternative methods of solving the partial differential equation of the base Hamiltonian in an effort to discover some simplifications.

---

*This research was supported by NASA Research Grant NGR-43-002-015 and was done under the direction of M. G. Boyce. A part of it was included in the author's master's thesis in mathematics at Vanderbilt University.

HAMILTON-JACOBI THEORY APPLICATIONS

A planar rocket trajectory problem is formulated with end conditions allowing for various missions, including rendezvous with a satellite in a coplanar orbit. The base Hamiltonian is taken as the part not involving thrust. The partial differential equation for it is linear, and our first solution uses Lagrange's method for obtaining a complete integral. Jacobi's equations determining original variables in terms of canonic constants are used to eliminate the original variables from the perturbing Hamiltonian, the canonic constants becoming new generalized coordinates and momenta.

The canonical equations of the new Hamiltonian are then the differential equations of the extremals.

The second method of solving the partial differential equations for the base Hamiltonian is to first transform it by a canonical transformation of variables and then use Jacobi's method to find a complete integral. The procedure described above is then repeated.

HAMILTON-JACOBI THEORY

Mayer Control Problem

The Mayer problem of calculus of variations involving control variables may be expressed in the following form.

The problem is to find in a class of admissible arcs

$$y_i(t),\ u_j(t),\ t_o \le t \le t_1,\ i = 1, \cdots, n,\ j = 1, \cdots, m,$$

satisfying differential equations and end conditions

$$\dot{y}_i = f_i(t,y,u),$$

$$J_k(t_o, y(t_o), t_1, y(t_1)) = 0, \qquad k = 1, \cdots, p \le 2n + 2,$$

one which will minimize a function

$$J(t_o, y(t_o), t_1, y(t_1)) .$$

Here, in the arguments of the functions, y denotes the n-vector $y_1, \cdots, y_n$ and u the m-vector $u_1, \cdots, u_m$. The super dot denotes derivative with respect to t. Partial derivatives will often be denoted by subscript variables and summation by the tensor analysis device of repeated indices. In this study admissible arcs will be arcs whose elements $(t, y, \dot{y})$ lie in a given $2n + 1$ dimensional open region R and whose control variables u are in an open region U. The end points

$(t_o, y(t_o), t_1, y(t_1))$ of admissible arcs are required to lie in an open set S, and y, $\dot{y}$, u are continuous functions of t. The given functions f_i, J_k, J are assumed to have continuous partial derivatives in their arguments to as high as second order.

First Necessary Condition: Multiplier Rule

The classical first necessary condition can be stated for the Mayer problem with control variables in the following form. [4]

Theorem 1. An admissible arc E is said to satisfy the multiplier rule if there exists a function

$$H(t,y,u,\lambda) = \lambda_i f_i, \qquad i = 1,2,\cdots,n,$$

where λ's are functions of t not simultaneously zero and continuous along the arc E, such that the equations

$$(1) \qquad \dot{\lambda}_i = H_{y_i}, \quad \dot{y}_i = -H_{\lambda_i}, \quad H_{u_j} = 0, \quad j = 1,\cdots,m,$$

are satisfied, if the end point conditions $J_k = 0$, $k = 1, \cdots, p$, hold, and if the transversality matrix

$$\begin{Vmatrix} H(t_o) + J_{t_o} & -H(t_1) + J_{t_1} & -\lambda_i(t_o) + J_{y_i}(t_o) & \lambda_i(t_1) + J_{y_i}(t_1) \\ J_{kt_o} & J_{kt_1} & J_{ky_i}(t_o) & J_{ky_i}(t_1) \end{Vmatrix}$$

is of rank p. Every minimizing arc must satisfy the multiplier rule.

Solutions of equations (1) are called extremals, and equations (1) are called the canonical equations of extremals. They are the Euler-Lagrange equations for the problem, and the function H is analogous to the Hamiltonian of mechanics. If, for admissible arcs, y and u are assumed only piecewise continuous, then Theorem 1 holds between corners of E.

Weierstrass Condition

The Weierstrass condition for the Mayer control problem can be stated as follows. [4]

Theorem 2. Along the minimizing arc E, the inequality

$$H(t,y,\lambda,\bar{u}) \leq H(t,y,\lambda,u)$$

must hold at each element (t,y,λ,u) of E for every $\bar{u}$ in U.

Thus $H(t,y,\lambda,u)$ is a maximum with respect to the control variables for a minimizing arc, for which reason this condition is often called the Maximum Principle.

Elimination of Control Variables

An arc along which the determinant $\left| H_{u_j u_h} \right| \neq 0$ is said to be non-singular. It will be assumed that all arcs considered are non-singular. The equations $H_{u_j} = 0$ can then be solved for the control variables in terms of multipliers and state variables, and control variables can be eliminated from the Hamiltonian. This will be supposed done, and the Hamiltonian will be written as

$$H^*(t,y,\lambda) = H(t,y,u(t,y,\lambda),\lambda).$$

It follows that the canonical equations of the extremals can be expressed in terms of H^*. For, if the equations

$$H_{u_j} = 0, \quad j = 1,\cdots,m,$$

of the set of equations (1) can be solved for $u_j = u_j(t,y,\lambda)$, then

$$H^*_{y_i} = H_{y_i} + H_{u_j} u_{jy_i},$$

$$H^*_{\lambda_i} = H_{\lambda_i} + H_{u_j} u_{j\lambda_i}.$$

Since $H_{u_j} = 0$, it follows that

$$H^*_{y_i} = H_{y_i} \qquad \text{and} \qquad H^*_{\lambda_i} = H_{\lambda_i}$$

or $$\dot{y}_i = H_{\lambda_i} = H^*_{\lambda_i}$$

and $$\dot{\lambda}_i = -H_{y_i} = -H^*_{y_i} \ .$$

Hereafter $H^*(t,y,\lambda)$ will be denoted by $H(t,y,\lambda)$ because of the equivalence of the two Hamiltonians.

The Hamilton-Jacobi Equation

The partial differential equation of first order

$$S_t + H(t,y,S_y) = 0, \tag{2}$$

is called the Hamilton-Jacobi equation. It has dependent variable S and $n + 1$ independent variables $t,y_1,\cdots,y_n$. The complete solution of (2) will have $n + 1$ arbitrary constants. However, one is additive and is of no importance here, so we shall consider a solution with n independent constants, no one of which is additive, to be a complete solution.

Theorem 3. Let the Hamilton-Jacobi equation (2) have the solution $S = S(t,y_1,\cdots,y_n,\alpha_1,\cdots,\alpha_m)$ depending on m $(\leq n)$ parameters $\alpha_1,\cdots,\alpha_m$. Then each derivative S_{α_j} is a first integral of the canonical Euler equations system

$$\dot{y}_i = H_{\lambda_i} \ , \quad \dot{\lambda}_i = -H_{y_i} \ ;$$

that is, S_{α_j} = constant along an extremal.

HAMILTON-JACOBI THEORY APPLICATIONS

Jacobi's Theorem

Theorem 4. Let $S(t,y_1,\cdots,y_n,\alpha_1,\cdots,\alpha_n)$ be a complete integral of the Hamilton-Jacobi equation (2), that is, a solution depending on n-parameters $\alpha_1,\cdots,\alpha_n$ and having the n by n determinant $\left|S_{\alpha_i y_h}\right| \neq 0$. Also let $\beta_1,\cdots,\beta_n$ be n arbitrary constants. Then the functions

$$(3) \qquad y_i = y_i(t,\alpha_1,\cdots,\alpha_n,\beta_1,\cdots,\beta_n), \qquad i = 1,\cdots,n$$

defined by the relations $S_{\alpha_i} = \beta_i$, together with the functions $\lambda_i = S_{y_i}$, constitute a general solution of the canonical system

$$\dot{y}_i = H_{\lambda_i}, \quad \dot{\lambda}_i = -H_{y_i}, \quad i = 1,\cdots,n.$$

For proofs of theorems 3 and 4 see [5, p. 90].

Hamilton-Jacobi Perturbation Theory

In celestial mechanics the path of a planet is disturbed by the presence of other heavenly bodies. This disturbing force is very small compared to the attraction of the sun. The Hamiltonian is expressed as a sum of two parts; the one which corresponds to the motion of the planet without the disturbing influence is called the base Hamiltonian, and the one corresponding to the disturbing factor is called the perturbing Hamiltonian. The low thrust rocket problems in trajectory analysis can be treated in a similar way, the thrust of the engine being considered as the disturbing factor.

The following theorem shows how to obtain a complete integral of order n of the Hamilton-Jacobi equation for the base Hamiltonian in case it involves fewer than n λ's [6, p. 29].

Theorem 5. Let $H(t,y_1,\cdots,y_n,\lambda_1,\cdots,\lambda_n)$ be the Hamiltonian for a dynamical system. Let $H_o = H_o(t,y_1,\cdots,y_n,\lambda_1,\cdots,\lambda_k)$, where $k < n$, be the base Hamiltonian and let $S^*(t,y_1,\cdots,y_n,\alpha_1,\cdots,\alpha_k)$ be a solution of the Hamilton-Jacobi equation for H_o depending on k independent parameters $(\alpha_1,\cdots,\alpha_k)$ with $\left|S^*_{y_i\alpha_j}\right| \neq 0, i,j = 1,2,\cdots,k$. Then

$$S = S^*(t,y_1,\cdots,y_n,\alpha_1,\cdots,\alpha_k) + \sum_{i=k+1}^{n} \alpha_i y_i ,$$

where $(\alpha_{k+1},\cdots,\alpha_n)$ are independent parameters, is a complete solution of order n for the base Hamilton-Jacobi equation.

From Theorem 4 it follows that

$$(4) \qquad \beta_i = S_{\alpha_i}, \qquad \lambda_i = S_{y_i}, \qquad i = 1,\cdots,n.$$

We solve these equations for y's and λ's in terms of α's and β's, thus $y_i = y_i(\alpha,\beta,t)$ and $\lambda_i = \lambda_i(\alpha,\beta,t)$, and substitute these values in the perturbing Hamiltonian, say H_1. Now H_1 is expressed in α's and β's as variables.

On considering S to be a generating function for a canonical transformation with α's and β's as new variables, it follows that the new Hamiltonian is $S_t + H$, [5, p. 79]. But

$$S_t + H = S_t + H_o + H_1, \quad \text{and} \quad S_t + H_o = 0$$

when S is a complete integral of the Hamilton-Jacobi equation for the base solution. Hence the H_1 is the Hamiltonian for the total problem in terms of the variables α_i, β_i ; and the canonical equations for extremals in these coordinates are

$$\dot{\alpha}_i = H_{1\beta_i}, \qquad \dot{\beta}_i = -H_{1\alpha_i} .$$

The solution of these equations gives the extremals for the problems with $2n$ constants of integration [6, p. 27; 7, p. 137]. By the use of the set of equations (4) we can express the trajectory in terms of y's and t. This theory can be extended to splitting the Hamiltonian into more than two parts.

Canonical Transformation

Suppose the variables y's and λ's are transformed to new variables q's and p's. If the transformation has the property that for every Hamiltonian $H(t,y,\lambda)$ there exists a function $K(t,q,p)$ such that

$$\dot{q} = \frac{\partial K}{\partial p_i} , \quad \dot{p} = \frac{-\partial K}{\partial q_i} \qquad i = 1,\cdots,n,$$

then the transformation is canonical.

It is assumed that the transformation has a non-vanishing Jacobian M. Let N denote the matrix $\begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}$ of order 2n, then the necessary and sufficient condition that the transformation be canonical is that $M^T N M = cN$, where c is a non-zero real number. [8].

PLANAR TRAJECTORY OPTIMIZATION PROBLEM

A rocket moving in a plane under a central gravitational force and the thrust of an engine is to achieve a specified mission starting from a given initial state. The variable angle of thrust, which is a function of time, is the control variable. It is desired to find the equations of the path requiring the least amount of fuel.

Assumptions

The path of the rocket is assumed to be in a plane, and hence a polar coordinate system is used, with origin at the center of the earth. The coordinate system is fixed relative to the earth and the gravitational force on the rocket is assumed directed towards the origin. The rocket is considered as a particle of variable mass. Air resistance is assumed negligible and thrust magnitude to be proportional to a constant rate of flow of mass.

Equations of Motion

Let (r,θ) be the polar coordinates of the rocket, α the angle between the radius vector and the direction of thrust, F the thrust magnitude, m the mass, c the constant rate of mass flow, and k the gravitational constant. The equations of motion can then be expressed as [9]

$$\begin{aligned} &\ddot{r} - r\dot{\theta}^2 = -k/r^2 + (F/m)\cos\alpha, \\ &r\ddot{\theta} + 2\dot{r}\dot{\theta} = (F/m)\sin\alpha, \\ &\dot{m} = -c. \end{aligned} \tag{5}$$

The theory of the Lagrangian

$$L = (\dot{r}^2 + r^2\dot{\theta}^2)/2 + k/r,$$

for the unit mass two-body problem without thrust, suggests defining

$$u = \partial L/\partial \dot{r}, \quad w = \partial L/\partial \dot{\theta}.$$

Thus $$u = \dot{r}, \quad w = r^2\dot{\theta},$$

and u is radial velocity, while w is rv, where v is tangential velocity. The equations of motion (5) then become

$$\begin{aligned} \dot{u} &= w^2/r^3 - k/r^2 + (F/m)\cos\alpha, \\ \dot{w} &= (r\,F/m)\sin\alpha, \\ \dot{r} &= u, \quad \dot{\theta} = w/r^2, \quad \dot{m} = -c\ . \end{aligned} \tag{6}$$

Let the initial and terminal conditions be denoted by $J_K = 0$, $K = 1,\cdots,p < 12$, in the notation of the general problem in the first part of this paper. For a minimum fuel trajectory, the function to be minimized can be expressed as $J \equiv m(t_o)$, with $m(t_1)$ a given constant. If the initial position and velocity are given, we have

$$J_1 \equiv t_o,\ J_2 \equiv u(t_o) - u_o,\ J_3 \equiv w(t_o) - w_o,\ J_4 \equiv r(t_o) - r_o,\ J_5 \equiv \theta(t_o) - \theta_o\ .$$

For terminal values, $J_6 = m(t_1) - m_1$, and the remaining J's would be functions of t_1, $u(t_1)$, $w(t_1)$, $r(t_1)$, $m(t_1)$.

Elimination of Control Variable by Weierstrass Condition

The Hamiltonian for equations (6) is

$$H = \lambda_1(w^2/r^3 - k/r^2 + (F/m)\cos\alpha) + \lambda_2(rF/m)\sin\alpha + \lambda_3 u + \lambda_4 w/r^2 - c\lambda_5\ ,$$

where the λ's are functions of t not simultaneously zero.

From the Weierstrass condition, H, as a function of α, must be a maximum. Hence $H_\alpha = -\lambda_1(F/m)\sin\alpha + \lambda_2 r(F/m)\cos\alpha = 0$ and

$$H_{\alpha\alpha} = -\lambda_1(F/m)\cos\alpha - \lambda_2 r(F/m)\sin\alpha \le 0\ .$$

It follows that

$$\tan\alpha = r\lambda_2/\lambda_1, \quad \sin\alpha = r\lambda_2/\sqrt{(\lambda_1^2 + r^2\lambda_2^2)}\ , \quad \cos\alpha = \lambda_1/\sqrt{(\lambda_1^2 + r^2\lambda_2^2)}$$

the radicals being positive because of $H_{\alpha\alpha} \le 0$.

Elimination of α gives $H = H_o + H_1$, where

$H_o = \lambda_1(w^2/r^3 - k/r^2) + \lambda_3 u + \lambda_4 w/r^2 - c\lambda_5$ and $H_1 = (F/m)\sqrt{(\lambda_1^2 + r^2\lambda_2^2)}$.

The Hamilton-Jacobi equation for the base Hamiltonian H_o then is

(7) $\partial S/\partial t + (w^2/r^3 - k/r^2)\,\partial S/\partial u + u\partial S/\partial r + (w/r^2)\,\partial S/\partial\theta - c\partial S/\partial m = 0.$

Determination of a Complete Integral

In seeking a complete integral, apply separation of variables, letting

$$S = S_1(t) + S_2(\theta) + S_3(m) + S_4(u,r).$$

The Hamilton-Jacobi equation assumes the form

$$dS_1/dt + (w^2/r^3 - k/r^2)\,\partial S_4/\partial u + u\partial S_4/\partial r - (w/r^2)\,dS_2/d\theta - cdS_3/dm = 0\ ,$$

which does not involve t, θ and m explicitly. Hence

$$dS_1/dt = \alpha_1\ ,\quad dS_2/d\theta = \alpha_2\ ,\quad dS_3/dm = \alpha_3\ ,$$

where α_1, α_2, α_3 are arbitrary constants.

The Hamilton-Jacobi equation can now be written as

(8) $(w^2/r^3 - k/r^2)\,\partial S_4/\partial u + u\partial S_4/\partial r = c\alpha_3 - \alpha_1 - \alpha_2 w/r^2\ ,$

which is in the form of Lagrange's linear equation [10, Ch.XII], and its subsidiary equations are

$$\frac{du}{w^2/r^3 - k/r^2} = \frac{dr}{u} = \frac{dS_4}{c\alpha_3 - \alpha_1 - \alpha_2 w/r^2}\ .$$

From the first subsidiary equation we get

(9) $u^2 - 2k/r + w^2/r^2 = -a^2,$

which we write as

$$f = -a^2\ ,$$

where $-a^2$ is a constant of integration with the sign chosen so as to give a periodic trajectory.

On substituting from the above for u in the last subsidiary equation, we have

$$dS_4 = \frac{((c\alpha_3 - \alpha_1)\,r^2 - \alpha_2 w)\,dr}{r\sqrt{-a^2r^2 + 2kr - w^2}}\ ,$$

the ambiguous sign of the radical being absorbed in the arbitrary constants $\alpha_1, \alpha_2, \alpha_3$. Integration now gives

$$S_4 = -\frac{A}{a^2}\sqrt{-a^2r^2 + 2kr - w^2} + \frac{Ak}{a^3}\sin^{-1}\frac{a^2r - k}{\sqrt{k^2 - a^2w^2}} - \alpha_2 \sin^{-1}\frac{kr - w^2}{r\sqrt{k^2 - a^2w^2}} + b,$$

where b is a constant of integration and $A = c\alpha_3 - \alpha_1$.

On eliminating a by use of (9) and introducing the abbreviating notations

$$X^2 = -w^2 + 2kr - u^2r^2,$$

$$Y^2 = (kr - w^2)^2 + u^2r^2w^2,$$

$$Z = kr - w^2,$$

we can express S_4 in the form

$$S_4 = \frac{Ar^3}{X^2}\left(-u + \frac{k}{X}\sin^{-1}\frac{X^2 - kr}{Y}\right) - \alpha_2 \sin^{-1}\frac{Z}{Y} + b,$$

or

$$S_4 = g + b.$$

The general solution of (8) will then be

$$\varphi(f, S_4 - g) = 0$$

where φ is an arbitrary differentiable function. It follows that

$$S_4 = g + \alpha_4 f + \alpha_6,$$

where α_4 and α_6 are arbitrary constants, is a solution and may be taken as a complete integral of (8). By adding S_1, S_2, S_3, S_4 we now obtain an integral of equation (7). As explained in the general discussion of Hamilton-Jacobi theory, the additive constant α_6 may be dropped. Also, by Theorem 5, the term $\alpha_5 w$ can be added to give, finally, as a complete integral of the Hamilton-Jacobi equation (7) for the base Hamiltonian H_o, the following

$$S = \alpha_1 t + \alpha_2 \theta + \alpha_3 m + \frac{Ar^3}{X^2}\left(-u + \frac{k}{X}\sin^{-1}\frac{X^2 - kr}{Y}\right) - \alpha_2 \sin^{-1}\frac{Z}{Y} - \frac{\alpha_4 X^2}{r^2} + \alpha_5 w.$$

The Remaining Canonic Constants

By Theorem 4 (Jacobi's Theorem), if S is a complete integral of the Hamilton-Jacobi equation, then there are constants $\beta_1, \cdots, \beta_5$ such that $\partial S/\partial \alpha_i = \beta_i$. On carrying out the differentiations on the above S, we get

$$\beta_1 = t - \frac{r^3}{X^2}\left(-u + \frac{k}{X}\sin^{-1}\frac{X^2 - kr}{Y}\right)$$

$$\beta_2 = \theta - \sin^{-1}\frac{Z}{Y} ,$$

$$\beta_3 = m + \frac{cr^3}{X^2}\left(-u + \frac{k}{X}\sin^{-1}\frac{X^2 - kr}{Y}\right) = m - c(\beta_1 - t)$$

$$\beta_4 = -X^2/r^2 ,$$

$$\beta_5 = w .$$

The Multipliers

Also by Jacobi's theorem the λ's are equal to the partial derivatives of S with respect to u, w, r, θ, m ; and the equations so obtained together with the above equations determine a ten-parameter family of solutions of the canonical equations for the base Hamiltonian H_o. On letting B denote

$$k^2 + \beta_4\beta_5^2 ,$$

we find, after some simplification, the following results:

$$\lambda_1 = [2r\beta_4 - \beta_5^2/r + (r\beta_5 + k)\, k\beta_5^2/rB]\, A/\beta_4^2 + (2\alpha_4 - 3A(t - \beta_1)/\beta_4)u + (kr - \beta_5^2)\alpha_2\beta_5/rB,$$

$$\lambda_2 = [ur - 3\beta_4(t - \beta_1)]\, A\beta_5/r^2\beta_4^2 + 2\alpha_4\beta_5/r^2 + \alpha_5 + [Ak\beta_5(r\beta_4 - k) + \alpha_2\beta_4^2(rk + \beta_5^2)]\, u/r\beta_4^2 B,$$

$$\lambda_4 = \alpha_4 , \quad \lambda_5 = \alpha_5 .$$

The multiplier λ_3 can be computed in the same way as λ_1 and λ_2, but it is not needed for H_1.

The above computed λ_1 and λ_2 are expressed as functions of α's, β's, u, and r. The variable u can be eliminated, since from $\beta_4 = -X^2/r^2$ and $X^2 = -\beta_5^2 + 2kr - u^2r^2$ we get

$$u^2r^2 = \beta_4 r^2 + 2kr - \beta_5^2 .$$

Also we have, from the β_3 equation,

$$m = \beta_3 - c(t - \beta_1) .$$

The H_1 Hamiltonian

The perturbing Hamiltonian $H_1 = (F/m)\sqrt{\lambda_1^2 + r^2\lambda_2^2}$ can now be expressed as a function of α's, β's, r, and t, and r is a function of β's and t by means of the equation for β_1. As explained in the first part of this paper, H_1 is now the Hamiltonian for the total problem in variables α_i, β_i, t. Consequently, the canonical equations for the extremals in these variables are

$$\dot{\alpha}_i = \partial H_1/\partial\beta_i , \quad \dot{\beta}_i = -\partial H_1/\partial\alpha_i .$$

More explicitly, letting $\lambda = \sqrt{\lambda_1^2 + \lambda_2^2}$,

$$\dot{\alpha}_1 = \frac{cF\lambda}{(\beta_3 - c(t - \beta_1))^2} + \frac{F(\lambda_1 \partial\lambda_1/\partial\beta_1 + r^2\lambda_2 \partial\lambda_2/\partial\beta_1 + r\lambda_2^2 \partial r/\partial\beta_1)}{\lambda(\beta_3 - c(t - \beta_1))} ,$$

$$\dot{\alpha}_3 = \frac{-F\lambda}{(\beta_3 - c(t - \beta_1))^2} + \frac{F(\lambda_1 \partial\lambda_1/\partial\beta_3 + r^2\lambda_2 \partial\lambda_2/\partial\beta_3 + r\lambda_2^2 \partial r/\partial\beta_3)}{\lambda(\beta_3 - c(t - \beta_1))} ,$$

$$\dot{\alpha}_i = \frac{F(\lambda_1 \partial\lambda_1/\partial\beta_i + r^2\lambda_2 \partial\lambda_2/\partial\beta_1 + r\lambda_2^2 \partial r/\partial\beta_i)}{\lambda(\beta_3 - c(t - \beta_1))} , \quad i = 2, 4, 5 ,$$

$$\dot{\beta}_i = \frac{-F(\lambda_1 \partial\lambda_1/\partial\alpha_i + r^2\lambda_2 \partial\lambda_2/\partial\alpha_i)}{\lambda(\beta_3 - c(t - \beta_1))} , \quad i = 1, 2, 3, 4, 5.$$

Since λ_1 and λ_2 are linear in the α's, the differentiations in the right members of the $\dot{\beta}$ equations are easily carried out. However, this is not possible for the $\dot{\alpha}$ equations.

The solution of the above system of differential equations gives the optimal trajectories of the rocket in terms of α's, β's, t and ten constants of integration. Closed form solutions do not seem possible, so approximation methods by some type of iteration on r seem necessary.

A SECOND METHOD FOR THE PLANAR PROBLEM

This method involves a canonical transformation of variables and leads to a complete integral of the base Hamiltonian. As before, the perturbing Hamiltonian, with the canonic constants as new variables, becomes the Hamiltonian for the total problem. The resulting canonical differential equations of extremals are somewhat different from those of our first method, but they again involve similar inherent difficulties and do not lead to closed form solutions.

The Canonical Transformation

Let the following transformation be made, where the q's denote the generalized coordinates and the p's the generalized momenta.

$$\lambda_1 = q_1 , \qquad -u = p_1 ,$$

$$w = q_2 , \qquad \lambda_2 = p_2 ,$$

$$r = q_3 , \qquad \lambda_3 = p_3 ,$$

$$\theta = q_4 , \qquad \lambda_4 = p_4 ,$$

$$\lambda_5 = q_5 , \qquad -m = p_5 .$$

This transformation is easily verified to satisfy the necessary and sufficient condition for a canonical transformation as given in the first part of this paper.

The other transformations consisting of interchange of coordinates and momenta have been investigated. Changing r to a momentum variable greatly complicates the Hamilton-Jacobi equation. Changing θ has little effect on either the base or the total Hamiltonian. Changing u only, or u and w , or u, w, and m to momenta give essentially the same S for the base Hamiltonian as does the above transformation.

A Complete Integral of the Base Hamiltonian

In the new variables, H_o assumes the following form:

$$H_o = q_1(q_2^2/q_3^3 - k/q_3^2) - p_1 p_3 + p_4 q_2/q_3^2 - cq_5 .$$

Hence the Hamilton-Jacobi equation is

$$(10) \quad S_t + q_1(q_2^2/q_3^3 - k/q_3^2) - p_1 p_3 + p_4 q_2/q_3^2 - cq_5 = 0,$$

where S_t, p_1, p_3, p_4 represent $\partial S/\partial t$, $\partial S/\partial q_1$, $\partial S/\partial q_3$, and $\partial S/\partial q_4$,

respectively. A solution of the above partial differential equation can be obtained by Jacobi's method [11].

$$\frac{dS_t}{0} = \frac{dp_4}{0} = \frac{dp_1}{-q_2^2/q_3^3 + k/q_3^2} = \frac{dq_3}{-p_1}$$

The first two terms give $S_t = \partial S/\partial t = \alpha_1$, $p_4 = \partial S/\partial q_4 = \alpha_4$, and the last two give

$$p_1 dp_1 = (q_2^2/q_3^3 - k/q_3^2)dq_3 .$$

On integrating this we get

$$p_1^2 = (-\alpha_3^2 q_3^2 + 2kq_3 - q_2^2)/q_3^2$$

or $p_1 = W/q_3$ where $W = \pm\sqrt{-\alpha_3^2 q_3^2 + 2kq_3 - q_2^2}$.

The constant of integration, $-\alpha_3^2$, has been chosen negative to give a periodic solution.

When the values for S_t, p_1, p_4 are substituted in the equation (10) and the result solved for p_3, we get

$$p_3 = \frac{\partial S}{\partial q_3} = \frac{(\alpha_1 - cq_5)q_3}{W} + \frac{\alpha_4 q_2 - kq_1}{q_3 W} + \frac{q_1 q_2^2}{q_3^2 W} .$$

Then

$$\int \frac{\partial S}{\partial q_3}\, dq_3 = \left(\frac{\alpha_1 - cq_5}{-\alpha_3^2} + \frac{q_1}{q_3}\right) W \pm \frac{k(\alpha_1 - cq_5)}{\alpha_3^3} \sin^{-1} \frac{\alpha_3^2 q_3 - k}{\sqrt{k^2 - \alpha_3^2 q_2^2}}$$

$$\pm \alpha_4 \sin^{-1} \frac{kq_3 - q_2^2}{q_3\sqrt{k^2 - \alpha_3^2 q_2^2}} .$$

The solution of the partial differential equation (10) is obtained from the exact differential

$$dS = (\partial S/\partial t)dt + (\partial S/\partial q_1)dq_1 + (\partial S/\partial q_3)dq_3 + (\partial S/\partial q_4)dq_4 .$$

However, q_2 and q_5 need to be included as independent variables in addition to t, q_1, q_3, and q_4; so, by use of Theorem 5, together with

the above results, we get

$$S = \alpha_1 t + \alpha_4 q_4 + \left(\frac{q_1}{q_3} - \frac{\alpha_1 - cq_5}{\alpha_3^2} \right) W \pm \frac{k(\alpha_1 - cq_5)}{\alpha_3^3} \sin^{-1} \frac{\alpha_3^2 q_3 - k}{\sqrt{k^2 - \alpha_3^2 q_2^2}}$$

$$\pm \alpha_4 \sin^{-1} \frac{kq_3 - q_2^2}{q_3 \sqrt{k^2 - \alpha_3^2 q_2^2}} + \alpha_2 q_2 + \alpha_5 q_5$$

as a complete integral of (10) involving five parametric constants, the additive constant being ignored. Note that the term $q_1 W/q_3$ occurs twice in the integrations but is counted only once in S.

The Canonical Constants β_i and Momenta p_i

By Jacobi's Theorem, $\partial S/\partial \alpha_i = \beta_i$, with arbitrary β_i. Let

$$C = \alpha_i - cq_5 , \qquad D = \sqrt{k^2 - \alpha_3^2 q_2^2} .$$

Then

$$\beta_1 = t - W/\alpha_3^2 \pm (k/\alpha_3^3) \sin^{-1}(\alpha_3^2 q_3 - k)/D,$$

$$\beta_2 = q_2 ,$$

$$\beta_3 = 3C(t - \beta_1)/\alpha_3 - CW/\alpha_3^3 - \alpha_3 q_1 q_3/W + q_3^2 C/\alpha_3 W + kq_3 C/\alpha_3^3 W$$
$$\mp (k^2 C + \alpha_3^4 \alpha_4 q_2)(kq_3 - q_2^2)/\alpha_3^3 WD^2,$$

$$\beta_4 = q_4 \pm \sin^{-1}(kq_3 - q_2^2)/q_3 D ,$$

$$\beta_5 = q_5 .$$

Where ambiguous signs occur above, the top sign is to be taken if W is chosen positive, otherwise the lower sign.

Since $p_i = \partial S/\partial q_i$, it follows that

$$p_1 = W/q_3 ,$$

$$p_2 = \alpha_2 + [C/\alpha_3^2 - \alpha_4/q_2 + (kC/\alpha_3^2 D^2 + \alpha_4 k/q_2 D^2)(\alpha_3^2 q_3 - k) - q_1/q_3](q_2/W),$$

$$p_3 = q_3 C/W + (\alpha_4 q_2 - kq_1)/q_3 W + q_1 q_2^2/q_3^2 W , \qquad p_4 = \alpha_4 ,$$

$$p_5 = \alpha_5 + cW/\alpha_3^2 \mp (kc/\alpha_3^3) \sin^{-1} (\alpha_3^2 q_3 - k)/D = \alpha_5 + c(t - \beta_1) .$$

The Hamiltonian H_1

Application of the canonical transformation to the original H_1 gives

$$H_1 = (-F/p_5) \sqrt{q_1^2 + p_2^2 q_3^2} .$$

To express H_1 in terms of α's, β's, t and q_3, we find from $\beta_2 = q_2$, $\beta_5 = q_5$, and the β_3 equation that

$$q_1 = [3C(t - \beta_1) - \alpha_3 \beta_3 - CW/\alpha_3^2 + q_3^2 C/W](W/\alpha_3^2 q_3) + [kq_3 CD^2 + (k^2 C + \alpha_3^4 \alpha_4 \beta_2)(kq_3 - \beta_2^2)]/\alpha_3^4 q_3 D^2 ,$$

where now $C = \alpha_1 - c\beta_5$, $D = \sqrt{k^2 - \alpha_3^2 \beta_2^2}$, $W = \pm \sqrt{-\beta_2^2 + 2kq_3 - \alpha_3^2 q_3^2}$.

This value of q_1 substituted in the formula for p_2 above gives

$$p_2 = \alpha_2 - \alpha_4/W + \beta_2 CW/\alpha_3^4 q_3^2 - k\beta_2 C/\alpha_3^4 q_3 W + [\alpha_3 \beta_2 \beta_3 - 3\beta_2 C(t - \beta_1)]/\alpha_3^2 q_3^2 + [\alpha_3^2 kq_3^2 (\beta_2 C + \alpha_3^2 \alpha_4)(\alpha_3^2 q_3 - k) + \beta_2 (k^2 C + \alpha_3^4 \alpha_4 \beta_2)(kq_3 - \beta_2^2)]/\alpha_3^4 q_3^2 D^2 W .$$

By using the expressions for p_2, q_1, and p_5 as above, we can

reduce H_1 to a function of α's, β's, q_3, and t. By the equation for β_1, q_3 is an implicit function of α's, β's, and t. Hence H_1 becomes a function of α's, β's, and t, and is then the Hamiltonian for the total problem. The canonical equations of extremals giving the optimal trajectory can then be obtained as in the first method. The analysis again is very involved and does not lead to closed form solutions, so we do not proceed further here. For another treatment of this problem one should refer to W. F. Powers [12], The search for canonical transformations which will give simpler forms for $\dot{\alpha}_i$ and $\dot{\beta}_i$ should be continued.

HAMILTON-JACOBI THEORY APPLICATIONS

REFERENCES

1. Hestenes, Magnus R. A General Problem in the Calculus of Variations with Application to Paths of Least Time. U. S. Air Force Project Rand Research Memorandum RM-100. Santa Monica, Calif.: Rand Corporation, March 1, 1950.

2. Hestenes, Magnus R. Calculus of Variations and Optimal Control Theory. New York: John Wiley and Sons, Inc., 1966.

3. Miner, William E. The Equations of Motion for Optimized Propelled Flight Expressed in Delannay and Poincare Variables and Modifications of These Variables. NASA Technical Note NASA TN D-4478. Washington, D. C.: National Aeronautics and Space Administration, May, 1968.

4. Boyce, M. G. and Linnstaedter, J. L. "Necessary Conditions for a Multistage Bolza-Mayer Problem Involving Control Variables and Having Inequality and Finite Equation Constraints," Progress Report No. 7 on Studies in the Fields of Space Flight and Guidance Theory. Huntsville, Alabama: NASA-MSFC, Aero-Astrodynamics Laboratory, 1965.

5. Gelfand, I. M. and Fomin, S. V. Calculus of Variations. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1963.

6. Powers, W. F. Hamiltonian Perturbation Theory for Optimal Trajectory Analysis. Austin, Texas: Engineering Mechanics Research Laboratory, University of Texas, 1966.

7. Smart, W. M. Celestial Mechanics. New York: John Wiley and Sons, Inc., 1961.

8. Powers, W. F. and Tapley, B. D. Canonical Transformation Theory and the Optimal Trajectory Problem. Austin, Texas: Engineering Mechancics Research Laboratory, University of Texas, August, 1967.

9. Loney, S. L. An Elementary Treatise on the Dynamics of a Particle and of Rigid Bodies. Cambridge, England: University Press, 1956.

10. Piaggio, H. T. H. An Elementary Treatise on Differential Equations and Their Applications. London: G. Bell and Sons, Ltd., 1928.

11. Miller, Frederic H. Partial Differential Equations. New York: John Wiley and Sons, Inc., 1960.

12. Powers, W. F. Canonical Transformation Theory and the Optimal Low-Thrust Problem. Austin, Texas: Engineering Mechanics Research Laboratory, University of Texas, March 1968.

ON A METHOD OF OBTAINING A COMPLETE INTEGRAL OF THE HAMILTON-JACOBI EQUATION ASSOCIATED WITH A DYNAMICAL SYSTEM

By Philip M. Fitzpatrick
Professor of Mathematics

and

John E. Cochran
Instructor of Aerospace Engineering
Auburn University
Auburn, Alabama

ON A METHOD OF OBTAINING A COMPLETE INTEGRAL OF THE HAMILTON-JACOBI EQUATION ASSOCIATED WITH A DYNAMICAL SYSTEM

By

Philip M. Fitzpatrick
Professor of Mathematics

and

John E. Cochran
Instructor of Aerospace Engineering
Auburn University, Auburn, Alabama

Consider a dynamical system whose equations of motion are

$$\left.\begin{aligned} \dot{q}_i &= \frac{\partial H(q_j;p_j;t)}{\partial p_i} \\ \dot{p}_i &= -\frac{\partial H(q_j;p_j;t)}{\partial q_i} \end{aligned}\right\} \quad i=1,2,\ldots,n;\ j=1,2,\ldots,n \tag{1}$$

where the Hamiltonian, $H(q_j;p_j;t)$, is understood to be a function of the generalized coordinates, q_j, and their conjugate momenta, p_j, $j=1,2,\ldots,n$, and possibly the time, t. If one-half of the integrals of Eqs (1) have been obtained in a suitable form, there is a well-known theorem, due to Liouville,[1] which may be used to find the remaining integrals. The purpose of this note is to point up the related, but perhaps not so well-known fact that a method of obtaining a complete integral of the Hamilton-Jacobi partial differential equation associated with (1) is implicitly contained in the theorem. Since a complete integral of (1) will permit us to express the solution of (1) in terms of canonical constants of integration, recognition of this fact is of importance in studying perturbations of the original system. The method will be discussed and applied in what follows.

Suppose that n integrals of a dynamical system with 2n degrees of freedom are known in the form

$$\Phi_i(q_j;p_j;t) = \alpha_i, \qquad i=1,2,\ldots,n;\ j=1,2,\ldots,n \tag{2}$$

where the α_i form a set of n independent constants of integration. If the Poisson bracket expression, (Φ_i,Φ_j), vanishes for each i and j and if the Φ_i are solvable for the p_i in the form

---

[1] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (New York: Cambridge University Press, 1959), pp. 323-325.

$$p_i = f_i(q_j;\alpha_j;t), \qquad i=1,2,\ldots,n;\ j=1,2,\ldots,n \tag{3}$$

the Liouville theorem states that the difference between

$$\sum_{i=1}^{n} f_i\,dq_i$$

and $H(q_j;\alpha_j;t)dt$ is the perfect differential of a function $W(q_j;\alpha_j;t)$ and that the remaining n integrals of the system are given by

$$\frac{\partial W}{\partial \alpha_i} = \beta_i, \qquad i=1,2,\ldots,n \tag{4}$$

where the β_i form a set of n constants of integration which are independent of each other and of the set formed by the α_i.

To say that

$$\sum_{i=1}^{n} f_i\,dq_i - H(q_j;\alpha_j;t)dt, \qquad j=1,2,\ldots,n \tag{5}$$

is the perfect differential of a function $W(q_j;\alpha_j;t)$ means that

$$\frac{\partial W}{\partial q_i} = f_i = p_i, \qquad i=1,2,\ldots,n \tag{6}$$

$$\frac{\partial W}{\partial t} = -H \tag{7}$$

Thus, implicit in the Liouville theorem is the fact that the function W is a complete integral of (7) which is the Hamilton-Jacobi partial differential equation associated with the system.

When the n integrals of (2) can be solved for the q_i instead of the p_i, i=1,2,...,n, the theorem may also be applied, if the canonical transformation

$$\left.\begin{aligned} Q_i &= p_i \\ P_i &= -q_i \end{aligned}\right\} \tag{8}$$

to new variables (Q_i,P_i) is first introduced. Even if we are not able to solve the n integrals (2) explicitly for the p_i, or for the q_i, a complete integral may still be obtained in certain important cases now to be discussed.

Suppose we are able to solve the integrals (2) explicitly for ℓ(ℓ < n) momenta and n-ℓ coordinates. Suppose further that, after reordering the subscripts, the expressions for the ℓ momenta and n-ℓ coordinates can be written

A COMPLETE INTEGRAL OF THE HAMILTON-JACOBI EQUATION

in the restricted form

$$\left.\begin{aligned} p_i &= f_i(q_k;p_m;\alpha_j;t), && i=1,2,\ldots,\ell;\ k\leq\ell;\\ & && m>\ell;\ j=1,2,\ldots,n\\ q_i &= h_i(q_m;p_k;\alpha_j;t), && i=\ell+1,\ell+2,\ldots,n;\ k>\ell;\\ & && m\leq\ell;\ j=1,2,\ldots,n \end{aligned}\right\} \quad (9)$$

By introducing the canonical transformation

$$\left.\begin{aligned} p_i^* &= p_i \quad q_i^* = q_i, && i=1,2,\ldots,\ell\\ p_i^* &= -q_i \quad q_i^* = p_i, && i=\ell+1,\ell+2,\ldots,n \end{aligned}\right\} \quad (10)$$

Eqs (9) may be written in the form

$$p_i^* = f_i^*(q_j^*;\alpha_j;t), \qquad i=1,2,\ldots,n;\ j=1,2,\ldots,n \quad (11)$$

Equations (11) are in the form of (3), and the theorem may be applied.

Example 1: Central Orbit in the Plane, Polar Coordinates

For a particle moving in a plane under a central force derivable from the potential $V(r)$, the Hamiltonian function is a constant α_1. If we designate by (p_r,p_θ), the momenta conjugate to the polar coordinates (r,θ), respectively, (see Figure 1), the system has the well-known integrals

$$p_\theta = \alpha_2, \text{ a constant} \quad (12)$$

$$p_r = \pm\sqrt{2[\alpha_1 + V(r)] - \frac{\alpha_2^2}{r^2}} \quad (13)$$

From (5), we write

$$dW = p_r dr + p_\theta d\theta - \alpha_1 dt \quad (14)$$

If r_o is chosen so that no new independent constant is introduced, the function

$$W = \int_{r_o}^{r} p_r dr + \alpha_2\theta - \alpha_1 t \quad (15)$$

obtained by integrating (14), satisfies (7). Also, W is a complete integral of (7) since it contains two non-additive independent constants α_1 and α_2.

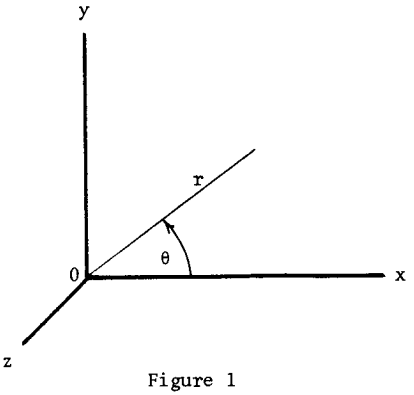


Figure 1

Example 2: Free Motion of a Triaxial Rigid Body

For the free rotations of a triaxial, rigid body about a fixed point 0, the Hamiltonian function, which is a constant of the motion, α_1, may be written in terms of the Euler angles (θ,ϕ,ψ), which specify the position of principal axes at 0 relative to space-fixed axes $0\xi\eta\zeta$ and their conjugate momenta (p_θ,p_ϕ,p_ψ). See Figure 2.

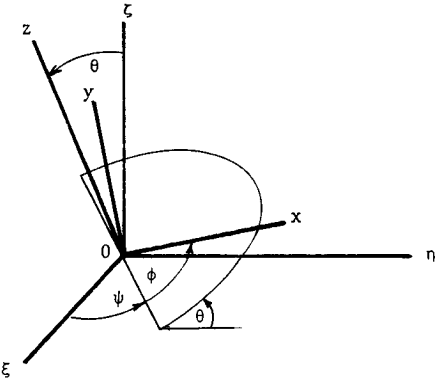


Figure 2

A COMPLETE INTEGRAL OF THE HAMILTON-JACOBI EQUATION

Three known integrals for this dynamical system are[2]

$$p_\psi = \alpha_3, \text{ a constant} \tag{16}$$

$$\theta = \tan^{-1}\left\{\frac{\sqrt{\alpha_2^2 - \alpha_3^2 - p_\theta^2}}{\alpha_3}\right\} - \tan^{-1}\left\{\frac{\sqrt{\alpha_2^2 - p_\phi^2 - p_\theta^2}}{p_\phi}\right\} \tag{17}$$

$$\phi = \tan^{-1}\left\{\frac{p_\theta}{\sqrt{\alpha_2^2 - p_\phi^2 - p_\theta^2}}\right\} + \tan^{-1}\left\{-\left(\frac{A}{B}\right)\frac{(2B\alpha_1 - \alpha_2^2)C + (C-B)p_\phi^2}{(2A\alpha_1 - \alpha_2^2)C + (C-A)p_\phi^2}\right\}^{\frac{1}{2}} \tag{18}$$

where A, B, and C are the principal moments of inertia at 0 and α_2 is the constant magnitude of the angular momentum about 0.

Although it is not possible to solve (17) and (18) so that p_ϕ and p_θ are expressed in the form of (3), the set of equations (16), (17), and (18) is of the form of (9); hence, the canonical transformation

$$\left.\begin{aligned} p_1 &= -\phi, & q_1 &= p_\phi \\ p_2 &= -\theta, & q_2 &= p_\theta \\ p_3 &= p_\psi, & q_3 &= \psi \end{aligned}\right\} \tag{19}$$

allows us to write (16), (17), and (18) in the form of (11). Then, from (5), we write

$$dW = p_1dq_1 + p_2dq_2 + p_3dq_3 - \alpha_1 dt \tag{20}$$

If q_{10} and q_{20} are chosen in a manner which introduces no new independent constants, the function

---

[2]See Whittaker, p. 325.

$$W = -\alpha_1 t + \alpha_3 q_3 + \int_{q_{20}}^{q_2} \tan^{-1}\left\{\frac{\sqrt{\alpha_2^2 - \alpha_3^2 - x^2}}{\alpha_3}\right\} dx$$

$$- \int_{q_{20}}^{q_2} \tan^{-1}\left\{\frac{\sqrt{\alpha_2^2 - q_1^2 - x^2}}{q_1}\right\} dx$$

$$+ \int_{q_{10}}^{q_1} \tan^{-1}\left\{-\left(\frac{A}{B}\right)\frac{(2B\alpha_1 - \alpha_2^2)C + (C - B)x^2}{(2A\alpha_1 - \alpha_2^2)C + (C - A)x^2}\right\}^{\frac{1}{2}} dx \qquad (21)$$

obtained by integrating (20), is a complete integral of (7).

AN OFFSET VECTOR ITERATION METHOD FOR SOLVING TWO-POINT BOUNDARY-VALUE PROBLEMS

By C. F. Price
Experimental Astronomy Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

AN OFFSET VECTOR ITERATION METHOD FOR SOLVING TWO-POINT BOUNDARY-VALUE PROBLEMS

By C. F. Price*

An offset vector iteration technique is proposed for solving two-point boundary-value problems. In this paper the properties of the method are explored. Application to parameter selection is first considered and convergence properties are described; comparison is made with other numerical methods. The two-point boundary-value problem is shown to be equivalent to the parameter selection problem. The method generally has a lower convergence rate than second order techniques; however, in many applications each iteration requires relatively few computational operations. Therefore it is competitive with higher order numerical procedures in applications that require few iterations to obtain an acceptably accurate solution. A modification to the offset vector method is suggested which takes advantage of the finite difference information generated at each iteration.

(First received September 1967 and in revised form February 1968)

1. Introduction

The use of offset vectors to develop iterative techniques for solving two-point boundary-value problems is a numerical procedure that has been proposed and investigated for use in near-earth (Godal, 1961), (Price and Boylan, 1964) and interplanetary guidance applications (Battin, 1964a), (Slater, 1966). The advantage of the method, when it can be applied, is that each iteration is often computationally simple to mechanise, relative to other techniques. In fact, there is evidence that it converges sufficiently rapidly in some cases to permit its use in real-time airborne guidance systems (Price *et al.*, 1964). This study was motivated by the desire to utilise an offset vector method for solving certain two-point boundary-value problems that represent necessary conditions for optimal trajectories. An example of such an application is presented in a recent paper (Price, 1967).

The concept of the offset vector method is easily understood and motivated through a simple, familiar example. Consider the problem of hitting a target with a projectile fired from a gun that is stationary with respect to the target. Let the direction of the gun barrel on the jth shot be designated by a unit vector, $i_j, j = 1, 2, \ldots$, expressed in an appropriate coordinate system. On the first shot, $j = 1$, i_1 is some function,

$$i_1 = i_1(r_T),$$

of the target's position, r_T. Suppose the first shot misses the target by a miss-vector, Δr_1, such that an impact point, r_1, is defined by

$$r_1 = r_T + \Delta r_1.$$

Using whatever quantitative knowledge of the miss he has, the gunner attempts to make an intelligent choice of the pointing direction on the next shot. If it happens that i_2 is expressed in the functional form (however crude)

$$i_2 = i_2(r_T - \Delta r_1)$$

where $(r_T - \Delta r_1)$ is a 'dummy' target position, we say that an offset vector iteration technique is being used. By analogy, on the kth iteration

$$i_{k+1} = i_{k+1}(r_T - \Delta r_1 - \Delta r_2 - \ldots - \Delta r_k);\ k = 1, 2\ldots$$

The philosophy is that on each iteration the aiming point is changed by the negative of the miss-vector. It is shown in this paper that such an approach is applicable to solving two-point boundary-value problems; in fact the above example can be formulated as such a problem.

Offset vector methods are *ad hoc* in nature because no general quantitative prescription is given for implementing the iterations. In the projectile example, the functional form of $i_{k+1}(\ )$ depends upon the sophistication of the fire control system. This point is emphasised in the subsequent discussion. However, it appears that the convergence properties of the technique can be described, to some extent, without reference to any special application, and comparisons can be made with other numerical procedures. That is the primary purpose of this paper.

In the next three Sections the concept of offset vectors for solving parameter selection problems is more precisely defined, convergence properties are described, and a simple example is presented. In Sections 5 and 6 it is shown that the two-point boundary-value problem reduces to that of parameter selection and results of utilising the method in a typical physical application are given. In Section 7 a modification to the offset vector method is suggested which takes advantage of the finite-difference information generated at each iteration. This provides a means for making a transition from the offset vector method to a finite-difference version of the

* Staff Member, *Experimental Astronomy Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts.*
This research has been sponsored by NASA ERC Contract No. NGR 22-009-207.

Newton–Raphson technique in situations where many iterations are required.

2. An offset vector method for solving the parameter selection problem

Parameter selection or equation solving is simply the task of finding a value of an n-dimensional vector $x = x_\infty$ which satisfies the vector equation

$$g(x) = 0. \tag{1}$$

In parameter optimisation problems, equations of this form are necessary conditions that a function $\phi(x)$ have a stationary point. We assume that $g(x)$ also has dimension n and that at least one solution of eqn. (1) exists.

Numerical techniques for solving eqn. (1) depend upon having an initial guess x_0 that is 'near' the desired solution x_∞ and improving that guess by iteratively generating a sequence $\{x_0, x_1, \ldots\}$ which converges to x_∞. Criteria for convergence of the sequence are usually given in terms of sufficient conditions satisfied by $g(x)$ in a region about x_∞ containing x_0.

The most important property of any particular numerical method is the total time required to achieve a sufficiently accurate solution for x_∞. This is dependent upon two factors—the number m of iterations required to obtain a value x_m that is sufficiently close to x_∞, and the computational complexity of each iteration. One often observes that these factors are inversely related; that is, the simpler each iteration is to perform, the more iterations required to obtain a desired level of accuracy in the solution. This characteristic is evidence of the fact that the amount of progress made in each iteration toward x_∞, i.e. the convergence rate, depends upon the amount of information used about $g(x)$ in deriving the recursion expressions.

Because the total time required for convergence is often dependent upon inversely related factors, it is difficult to state *a priori* in any particular application which of the various numerical methods is most advantageous from a computational point of view. However, if any initial guess x_0 is quite close to x_∞, relatively simple iteration techniques may accomplish the required degree of accuracy with no more, or few more, iterations than more elaborate methods. This rationale provides the motivation for describing an offset vector iteration technique which is potentially simple to implement and is based upon the idea of having a reasonably accurate initial guess x_0; in fact, the structure of the method is defined by the manner in which x_0 is chosen.

Suppose one can find an n-dimensional vector function $\hat{g}(x)$ that approximates $g(x)$ such that the solution $x \equiv x_0$ of

$$\hat{g}(x) = 0 \tag{2}$$

is relatively easily determined.* For example, $g(x)$ and

* This is not to say that x_0 need be determined by an explicit formula; the solution to eqn. (2) may also have to be obtained numerically. An example of this kind is given in Section 6.

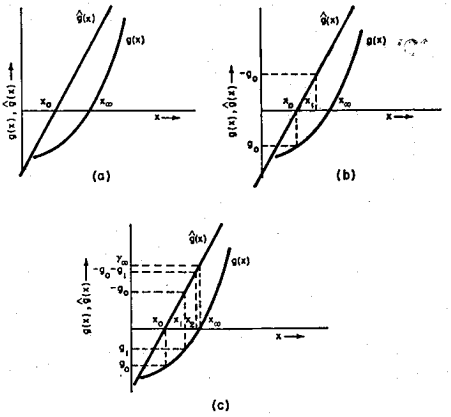


Fig. 1. Graphical development of the first two iterations of the offset vector method applied to a scalar function $g(x)$

$\hat{g}(x)$ may be of the form

$$\left.\begin{aligned} g(x) &= g + Gx + \epsilon f(x) = 0 \\ \hat{g}(x) &= g + Gx \qquad\quad = 0 \end{aligned}\right\} \tag{3}$$

with ϵ a constant scalar, g a constant vector, G a non-singular matrix and $f(x)$ some nonlinear function of x. If the term $\epsilon f(x)$ is small relative to $g(x)$ for x near x_∞, the solution $x_0 = -G^{-1}g$, is near x_∞. Let us write the solution to eqn. (2) as

$$x_0 = \hat{g}^{-1}(0) \equiv h(0) \tag{4}$$

where $\hat{g}^{-1}(\ )$ represents the required inversion of $\hat{g}(\ )$, and the argument 0 refers to the value of the right-hand side of eqn. (2). The situation is illustrated graphically in **Fig. 1a** for $n = 1$.

Having x_0, we can evaluate

$$g(x_0) \equiv g_0, \tag{5}$$

noting the eqn. (1) is in general not satisfied, that is, $g_0 \neq 0$. Based upon this observation an improvement to x_0 can be determined by the following reasoning. Suppose $\hat{g}(x)$ differs from $g(x)$ by only a constant vector f_0, that is,

$$\hat{g}(x) = g(x) + f_0; \quad \text{for all } x. \tag{6}$$

Then

$$g(x_0) = -f_0 = g_0.$$

If this be true, the solution to eqn. (1) is also the solution to

$$\hat{g}(x) - f_0 = 0$$

or

$$\hat{g}(x) = -g_0. \tag{7}$$

Thus we offset the approximating function by the negative of the error determined in eqn. (5) and calculate

x_1 from eqn. (7), using the notation of eqn. (4).

$$x_1 = h(-g_0). \tag{8}$$

This sequence of operations is illustrated in **Fig. 1b.** The quantity $-g_0$ is analogous to $-\Delta r_1$ in the projectile problem of the previous section.

In general, x_1 does not satisfy eqn. (6) either, as evidenced by

$$g(x_1) \equiv g_1 \neq 0.$$

Accordingly, replace eqn. (6) by the conjecture

$$\hat{g}(x) = g(x) - g_0 + f_1 \tag{9}$$

which leads to

$$g(x_1) = -f_1 = g_1$$
$$\hat{g}(x) = -g_0 - g_1 \tag{10}$$

resulting in

$$x_2 = h(-g_0 - g_1). \tag{11}$$

These steps are shown in **Fig. 1c.**

The recursion relationships required for the continuation of this method are readily inferred from the preceding discussion. Define

$$g_j = g(x_j)$$
$$\gamma_i = -\sum_{j=-1}^{i} g_j; \quad i = -1, 0, 1, \ldots \tag{12}$$
$$\gamma_{-1} = -g_{-1} = 0$$

and let

$$\hat{g}(x_i) = \gamma_{i-1}; \quad i = 0, 1, \ldots \tag{13}$$

Then,

$$\left.\begin{aligned} \gamma_i &= \gamma_{i-1} - g_i; \quad i = 0, 1, \ldots \\ x_i &= \hat{g}^{-1}(\gamma_{i-1}) = h(\gamma_{i-1}) \end{aligned}\right\} \tag{14}$$

At each iteration one evaluation each of $g(\ )$ and $h(\ )$ is required. The quantity γ_i is referred to as the *offset vector*. Now we shall discuss circumstances in which the sequence $\{x_0, x_1, \ldots\}$ generated by eqns. (12–14) converges to x_∞.

3. Convergence properties

One expects that the convergence properties of the offset vector method depend upon the accuracy with which $\hat{g}(x)$ approximates $g(x)$. To pursue this reasoning define an error vector $\Delta g(x)$ by

$$g(x) = \hat{g}(x) + \Delta g(x). \tag{15}$$

Substituting x_i for x, we have

$$g(x_i) = \hat{g}(x_i) + \Delta g(x_i). \tag{16}$$

Into eqn. (16) we can substitute for $\hat{g}(x_i)$ and x_i from eqns. (13) and (14), producing

$$g(x_i) = g_i = \gamma_{i-1} + \Delta g[h(\gamma_{i-1})]. \tag{17}$$

Rearranging terms and substituting for the quantity $(\gamma_{i-1} - g_i)$ from eqn. (14) yields

$$\gamma_i = -\Delta g[h(\gamma_{i-1})]. \tag{18}$$

Equation (18) is equivalent to eqn. (14) and is the recursion for solving

$$\gamma = -\Delta g[h(\gamma)]. \tag{19}$$

by successive approximations. The solution, γ_∞, to eqn. (19) is the limit of the sequence of offset vectors $\{\gamma_0, \gamma_1, \ldots\}$. Viewed another way, it is the value of $\hat{g}(x_\infty)$. (See Fig. 1.)

Sufficient conditions for the convergence of the sequence $\{\gamma_i\}$ are known for successive approximation iteration methods. For example, convergence is assured (Todd, 1962) if $\Delta g[h(\ )]$ satisfies the Lipschitz condition

$$\max|\Delta g[h(\gamma')] - \Delta g[h(\gamma'')]| < k \max|\gamma' - \gamma''|; \quad 0 < k < 1 \tag{20}$$

for all γ' and γ'' in a neighbourhood of γ_∞ containing $\gamma_{-1} = 0$.

Alternatively, a recursion relationship for x_i can be derived from eqn. (14). Substituting for γ_{i-1} and γ_{i-2} from respectively eqns. (14) and (13), we have

$$x_i = h[-\Delta g(x_{i-1})]. \tag{21}$$

The solution of this expression with x_i and x_{i-1} replaced by x is the value of $x = x_\infty$ that renders $g(x_\infty) = 0$ and $\hat{g}(x_\infty) = \gamma_\infty$.

A third way of viewing the iterative procedure is that the sequence $\{g_0, g_1, \ldots\}$ of evaluations of $g(x_i)$ is being driven to a limit of zero. This is perhaps the most natural point of view for the applications to be considered subsequently. From eqns. (12)–(14) it is evident that $g(x_i)$ is a nonlinear function of all $g(x_j)$, $j < i$, of the form

$$g_i = g[h(0 - g_0 - g_1 - \ldots - g_{i-1})]. \tag{22}$$

Similarly,

$$g_{i+1} = g[h(0 - g_0 - g_1 - \ldots - g_i)]. \tag{23}$$

Linearising g_{i+1} about g_i with substitution from eqns. (12)–(14) we have

$$g_{i+1} \cong g_i - G(x_i)H(\gamma_{i-1})g_i; \quad i = 0, 1, \ldots \tag{24}$$

where

$$G(x) \equiv \frac{\partial g(x)}{\partial x}; \quad H(\gamma) \equiv \frac{\partial h(\gamma)}{\partial \gamma}. \tag{25}$$

Equation (24) indicates that

$$\lim_{i \to \infty} g_i = 0$$

if

$$||I - GH|| < 1 \tag{26}$$

in some sufficiently small region about x_∞ such that the linearisation is valid. Note that if $g(\ ) = \hat{g}(\ )$, $GH = I$.

These convergence properties provide a comparison

between the offset vector method and other procedures that can be employed for finding x_∞. Considering eqn. (19), perhaps the most significant observation is that the method does not possess second-order convergence because the gradient matrix corresponding to eqn. (19),

$$\left.\frac{\partial \Delta g[h(\gamma)]}{\partial \gamma}\right|_{\gamma_\infty} \neq 0,$$

n general (Todd, 1962). Thus a Newton–Raphson technique, beginning at x_0 may require fewer iterations to approach x_∞ within a desired accuracy. However, the offset vector method possesses two advantages that motivate its use in certain situations.

First, applications arise in which $g(x)$ cannot be expressed in closed form, such as the solutions of many two-point boundary value problems. In these cases every evaluation of $g(x)$ requires numerical integration of differential equations. In addition, for Newton–Raphson-type procedures the gradient matrix must also be computed numerically, requiring additional complete integrations of the appropriate differential equations for each iteration. Hence, if the approximation $\hat{g}(x)$ is sufficiently accurate, one may conceivably reach a point sufficiently close to x_∞ with an offset vector technique before a higher order method gets started. The offset vector method has proved sufficiently rapid in situations of this kind to be incorporated in a real-time airborne guidance system (Price *et al.*, 1964). An example of such an application is included in Section 6.

Second, the offset vector method is a reasonable starting procedure for a higher order method in situations where many iterations are required. The points $x_0, x_1, \ldots$ and associated values $g_0, g_1, \ldots$ can be stored to provide corrections, based on finite differences, to subsequent evaluations of x_i. A possible method for accomplishing this is described in Section 7.

There is the disadvantage that some means must exist for finding an appropriate $\hat{g}(x)$. Whether this can be done depends upon the particular problem and the analyst's ingenuity; for this reason the concept of offset vectors does not provide a ready-made numerical algorithm for attacking all parameter selection problems. The fact that applications are known (see the references mentioned in Section 1 and the example of Section 6) where the method can be applied is a testimonial to its usefulness.

4. Example 1

To illustrate the offset vector method, a simple one-dimensional example is presented using equation numbers corresponding to those expressions in preceding sections which are exemplified.

Given

$$g(x) = 1 + x + \epsilon x^3 = 0. \tag{1}$$

Let

$$\hat{g}(x) = 1 + x. \tag{2}$$

Then

$$\gamma_i = \gamma_{i-1} - g_i$$
$$x_i = h(\gamma_{i-1}) = \gamma_{i-1} - 1. \tag{14}$$

Using the criterion for convergence provided by eqn. (26), we find that

$$G(x) = 1 + 3\epsilon x^2; \quad H(\gamma) = 1 \tag{25}$$
$$3x^2|\epsilon| < 1. \tag{26}$$

Furthermore, from eqn. (24)

$$\left|\frac{g_{i+1}}{g_i}\right| \cong 3x_i^2|\epsilon| \tag{24}$$

which provides a measure of the convergence rate.

It should be emphasised again that the offset vector method is not promoted especially for a high convergence rate. In general, and for this example in particular, it converges more slowly than Newton's method. The main advantage is the relative simplicity with which each iteration can be performed. This is illustrated by observing that the recursion relationships in eqn. (14) for this example require two subtractions and one evaluation of $g(x)$ per iteration. On the other hand, Newton's formula,

$$x_{i+1} = x_i - \frac{g(x_i)}{g'(x_i)}; \quad g'(x_i) = \left.\frac{dg(x)}{dx}\right|_{x=x_i}$$

requires one subtraction, one division, one evaluation of $g(x)$, and one evaluation of $dg(x)/dx$ per iteration; clearly this entails significantly more computation. The total time required to obtain an acceptably accurate solution for x_∞ is less for the offset vector method if $|\epsilon|$ is sufficiently small so that only one iteration of either method is required.

In situations where $g(x)$ has several dimensions and a complicated functional form, the computational advantages offered by an offset vector method are more significant. As mentioned previously, it is competitive with higher order techniques when a sufficiently good approximate solution can be obtained. In applications where the problem must be solved repeatedly, as in rocket guidance systems, considerable computational saving may be gained. This is illustrated by the example in Section 6.

5. The two-point boundary-value problem

The use of offset vectors to develop iterative techniques for solving two-point boundary-value problems is a numerical procedure that has been applied to near-earth (Godal, 1961), (Price *et al.*, 1964) and interplanetary guidance (Battin, 1964a), (Slater *et al.*, 1966) problems. In this section it is shown that the convergence properties can be stated in the same terms as for the parameter selection problem.

A two-point boundary-value problem is posed by assuming a given dynamical system described by n-dimensional vector differential equations

$$\dot{x} = f(x, t) \tag{27}$$

with prescribed end conditions

$$\left.\begin{aligned}\Phi[x(t_0), t_0] &= 0\\ \psi[x(t_f), t_f] &= 0\end{aligned}\right\} \tag{28}$$

where t_0 and t_f are initial and final times, x is an n-dimensional state vector, Φ and ψ are respectively l- and m-dimensional vectors, with $l + m = n + 2$. It is assumed that a solution exists which cannot be determined in closed form, requiring the use of numerical techniques.

We shall regard the solution to eqn. (27) known when the complete set of initial conditions $x(t_0)$, t_0 is determined such that eqns. (27) and (28) are satisfied. The explicit dependence upon eqn. (27) is conceptually eliminated by writing the solution as

$$x(t) = x[x(t_0), t_0, t] \tag{29}$$

so that eqn. (28) becomes

$$g[x(t_0), t_0, t_f] = \begin{bmatrix}\Phi[x(t_0), t_0]\\ \psi[x[x(t_0), t_0, t_f], t_f]\end{bmatrix} = 0 \tag{30}$$

Equation (30) has the form of eqn. (1) where the parameters to be determined are $x(t_0)$, t_0, and t_f.

The offset vector method is implemented in a manner analogous to that described in Section 2. Approximate solvable relations

$$\hat{g}[x(t_0), t_0, t_f] = 0 \tag{31}$$

are derived, often by means of a simplified set of differential equations

$$\dot{x} = \hat{f}(x, t), \tag{32}$$

subject to eqn. (28). For example, eqn. (27) may describe motion in a many-body gravitational field and eqn. (32) may represent an approximating two-body model with eqn. (28) specifying the initial and final positions at specified times. The solutions $x_0(t_{0_0})$, t_{0_0}, and t_{f_0} of eqn. (31) are entered as initial conditions into eqn. (27), and the differential equations are integrated from t_{0_0} to t_{f_0} producing

$$x_0(t_{f_0}) \equiv x[x_0(t_{0_0}), t_{0_0}, t_{f_0}]. \tag{33}$$

Substitution of t_{0_0}, t_{f_0} and $x_0(t_{0_0})$ for t_0, t_f, and $x(t_0)$ in eqn. (30) yields

$$g[x_0(t_{0_0}), t_{0_0}, t_{f_0}] = g_0 \neq 0$$

in general. Defining the vector

$$z^T = [x(t_0)^T, t_0, t_f],$$

the iterative computation of the sequence $\{z_0, z_1, \ldots\}$ proceeds just as in Section 2 with the understanding that each evaluation of

$$g[x_i(t_{0_i}), t_{0_i}, t_{f_i}] = g_i$$

requires integration of eqn. (27).

The motivation for using offset vectors is now more apparent. Vis-à-vis higher order methods it may be of considerable computational advantage to obtain even an algebraically complex form of eqn. (31) if computation of the gradient of $g[x(t_0), t_0, t_f]$ is thereby avoided. A practical multidimensional example of this type is considered in the next section. Observe that the projectile problem discussed in the Introduction can also be formulated as a two-point boundary-value problem and its solution obtained in the manner described above.

6. Example 2

This section discusses an application of the offset vector method to a practical two-point boundary-value problem. Equation numbers denote those expressions in previous sections which are exemplified.

Consider the motion of a body in a planar orbit in the earth's gravitational field. If the earth's rotation and atmospheric friction are neglected,* the equations of motion are reasonably accurately represented by

$$\left.\begin{aligned}\dot{x} &= v_x\\ \dot{v}_x &= -\frac{x}{r^3}\left[E + \frac{JEA^2}{r^2} - \frac{5JEA^2z^2}{r^4}\right]\\ \dot{z} &= v_z\\ \dot{v}_z &= -\frac{z}{r^3}\left[E + \frac{JEA^2}{r^2} - \frac{5JEA^2z^2}{r^4}\right] - \frac{2JEA^2z}{r^5}\end{aligned}\right\} \tag{27}$$

where A is the equatorial radius, J and E are constants, $r = \sqrt{(x^2 + z^2)}$, and x and z are position coordinates in an orthogonal coordinate system with the z axis along the earth's polar axis. Because the orbit is polar, only two dimensions need be considered. Equations (27) describe the gravitational accelerations including the effects of the earth's slightly elliptical shape. Let us pose the problem of finding the initial velocity components, $v_x(t_0)$ and $v_z(t_0)$, required to transfer a body from a given initial position at time $t_0 = 0$ to a given final position at a specified final time. Hence

$$\left.\begin{aligned}t_0 &= 0 & t_f - T_f &= 0\\ x(t_0) - a_x &= 0 & x(t_f) - b_x &= 0\\ z(t_0) - a_z &= 0 & z(t_f) - b_z &= 0\end{aligned}\right\} \tag{28}$$

where a_x, a_z, b_x, b_z, and T_f are given.

For the case where the earth's oblate effects are neglected ($J = 0$ in eqn. (27)), the task of finding the initial velocities subject to the given conditions is the familiar Lambert's problem of classical mechanics. For this case eqn. (27) can be integrated analytically by changing the independent variable; several methods of obtaining explicit expressions for $g(x)$ are known (Battin, 1964b). For $J \neq 0$, there is no known method of integrating eqn. (27) analytically; hence a numerical technique is required.

The offset vector method is naturally adapted to this application by using the known solution to Lambert's

* It is recognised that neglect of the earth's rotation contradicts the intent of treating a practical example. However, this effect can be included without changing the qualitative interpretation of the numerical results; it is omitted only to reduce the complexity of the discussion.

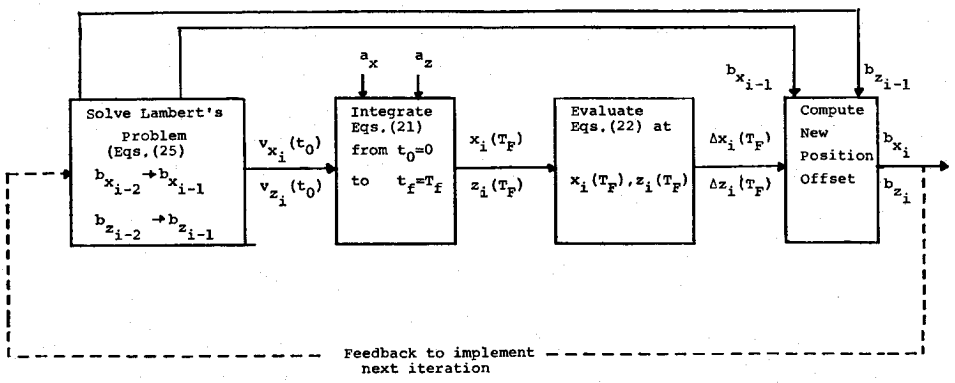


Fig. 2. Computational flow diagram of the *i*th iteration in Example 2

problem with $J = 0$ as an approximation. Introducing $J = 0$ into eqn. (27) produces a set of equations represented by eqn. (32) in Section 5. For the terminal conditions prescribed by eqn. (28), one form of $\hat{g}(x)$ due to Godal (Battin, 1964b) is given by

$$\left.\begin{aligned}
&v_x(0) - C_1(b_x - C_2 a_x) = 0\\
&v_z(0) - C_1(b_z - C_2 a_z) = 0\\
&C_1 - \frac{\sqrt{(EP)}}{r_f r_0 \sin\theta} = 0\\
&C_2 - 1 + \frac{r_f}{P}(1 - \cos\theta) = 0\\
&r_0 - \sqrt{(a_x^2 + a_z^2)} = 0\\
&r_f - \sqrt{(b_x^2 + b_z^2)} = 0\\
&\theta - \cos^{-1}[(a_x b_x + a_z b_z)/r_0 r_f] = 0\\
&P - \frac{\sqrt{(r_0 r_f)}\sin^2 0{\cdot}5\theta}{(B - \cos\alpha)\cos 0{\cdot}5\theta} = 0\\
&B - (r_0 + r_f)/2\sqrt{(r_0 r_f)}\cos 0{\cdot}5\theta = 0\\
&T_f - 2\left\{(\sqrt{(r_0 r_f)}\cos 0{\cdot}5\theta)^{1{\cdot}5}\sqrt{\left(\frac{B - \cos\alpha}{E}\right)}\right.\\
&\qquad\left.\left[1 + \frac{(B - \cos\alpha)(2\alpha - \sin 2\alpha)}{2\sin^3\alpha}\right]\right\} = 0
\end{aligned}\right\} \quad (31)$$

The solutions to eqns. (31) are the proper initial velocities to achieve the conditions in eqns. (28), neglecting the oblateness of the earth. Observe that eqns. (31) are transcendental in α; therefore their solution must be obtained numerically. This represents a situation where eqns. (2) cannot be inverted analytically.

The offset vector method proceeds by carrying out the following steps:

1. Denote the solutions of eqn. (31) as $v_{x0}(t_0)$ and $v_{z0}(t_0)$; these are obtained by any convenient numerical method. Newton's method has been used in this simulation.
2. Integrate eqn. (27) from $t = 0$ to $t = T_f$ using a_x, a_z, $v_{x0}(t_0)$ and $v_{z0}(t_0)$ as initial conditions. Denote position on this trajectory by $x_0(t)$ and $z_0(t)$.
3. Evaluate the left hand sides of eqn. (28) for the integrated trajectory. Define
$$\Delta x_0(T_f) \equiv x_0(T_f) - b_x$$
$$\Delta z_0(T_f) \equiv z_0(T_f) - b_z.$$
4. Recompute the initial velocities from eqn. (31) by requiring
$$x(T_f) - b_x = -\Delta x_0(T_f)$$
$$z(T_f) - b_z = -\Delta z_0(T_f).$$
This implies that eqn. (31) undergoes the changes of variable,
$$b_x \to b_x - \Delta x_0(T_f) \equiv b_{x_0}$$
$$b_z \to b_z - \Delta z_0(T_f) \equiv b_{z_0}.$$
Denote the solutions as $v_{x1}(t_0)$ and $v_{z1}(t_0)$.
5. Repeat steps 2 through 4 in an iterative fashion. The functional diagram in **Fig. 2** illustrates the steps at the *i*th iteration.

For this simulation the following parameter values are used:

$a_x = 2{\cdot}093 \times 10^7$ feet	$E = 1{\cdot}407645 \times 10^{16}$
$a_z = 0{\cdot}0$ feet	$J = 1{\cdot}62345 \times 10^{3}$
$b_x = 0{\cdot}0$ feet	$A = 2{\cdot}093 \times 10^7$ feet
$b_z = 3{\cdot}0 \times 10^7$ feet	$T_f = 2400{\cdot}0$ seconds

This roughly represents insertion into a 2000-mile altitude orbit at a point above the pole from a point on the equator. The computation was performed in double precision arithmetic on an IBM 360/65 computer. Newton's method is applied to solve Lambert's problem and a Gill-modified Runge–Kutta integration technique is used to integrate eqn. (27) with a 20 second time step. The values of terminal position error, $\Delta x_i(T_f)$ and $\Delta z_i(T_f)$, for two iterations are given in **Table 1**:

Table 1

Position error data from simulation

(Rounded off to 3 significant figures)

ERROR QUANTITY	$i = 0$ ERROR WITH NO CORRECTIONS TO LAMBERT SOLUTION	ERROR AFTER ith ITERATION	
		$i = 1$	$i = 2$
	FEET	FEET	FEET
$\lvert\Delta x_i(T_f)\rvert$	$2{\cdot}34 \times 10^4$	$1{\cdot}51 \times 10^1$	$4{\cdot}41 \times 10^{-2}$
$\lvert\Delta z_i(T_f)\rvert$	$5{\cdot}37 \times 10^4$	$7{\cdot}25 \times 10^1$	$9{\cdot}56 \times 10^{-2}$

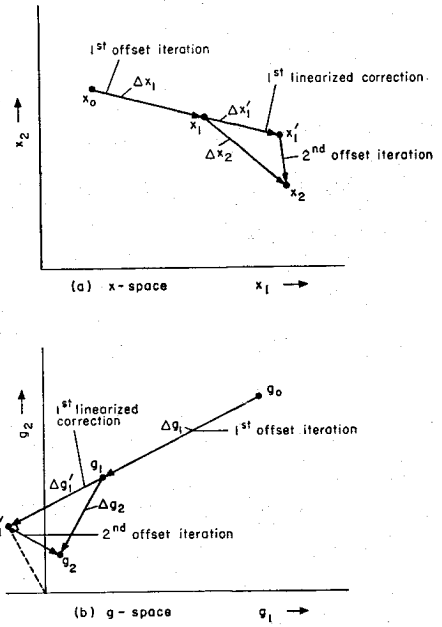


Fig. 3. Progress of finite-difference modification of offset vector method in two dimensions

Adequate accuracy is obtained in one iteration for many applications. For these cases any other numerical method that has an equal or greater convergence rate can be compared on the basis of the computational complexity of each iteration.

In this simulation the time required to solve Lambert's problem with sufficient accuracy is approximately 0·01 seconds whereas that required for integrating eqn. (27) is 0·30 seconds. Because the latter* dominates the former, any method that requires more differential equations to be integrated is at a competitive disadvantage with the offset vector method. For Newton–Raphson type procedures, the gradient matrix of $g(\ )$ with respect to $v_{xi}(t_0)$ and $v_{zi}(t_0)$ must be obtained. This can be obtained numerically by perturbing each velocity component separately and integrating eqn. (27) to determine the effect on the end conditions. Obtaining the complete gradient matrix by this procedure requires n additional complete integrations of eqn. (27) per iteration; this results in tripling the amount of integration required in this example, effectively tripling the computation time for each iteration. The gradient matrix can also be obtained by integrating the linear variational equations associated with eqn. (27); however, the increased computation is of the same order as that required to obtain the matrix by the perturbation technique.

These comparisons indicate that the offset vector method is superior to higher order methods in some problems. The example considered here has application to rocket guidance for which the thrust is directed so that the vehicle's velocity matches the values of $v_{xi}(t_0)$ and $v_{zi}(t_0)$ in Fig. 2. The two-point boundary-value problem must be solved many times in rapid succession because the initial time and the rocket's position are constantly changing. For 'real-time' computation of this sort, speed is a primary consideration.

* An integration step three or four times larger than 20 seconds would give terminal position accuracy better than 100 feet in this example.

7. Modified offset vector method

In Section 3 it is pointed out that the offset vector method can serve as a starting procedure for higher-order techniques. The possibility for doing this is evident at the $(n + 1)$th step after the sequences $\{x_0, x_1, \ldots x_n\}$ and $\{g_0, g_1, \ldots g_n\}$ have been computed. Defining

$$\left.\begin{aligned}\Delta x_i &\equiv x_i - x_{i-1}\\ \Delta g_i &\equiv g_i - g_{i-1}\end{aligned}\right\} \tag{34}$$

we have sufficient information to derive an approximate gradient matrix (or its inverse) provided the Δx_i's (or Δg_i's) are independent. For example,

$$\frac{\partial g}{\partial x} \cong \tilde{G} = X^{-1}\mathscr{G} \tag{35}$$

where $\mathscr{G}$ and X are matrices whose ith columns are respectively Δg_i and Δx_i. Faster convergence may possibly be obtained by continuing the numerical procedure with a Newton–Raphson-like technique using $\tilde{G}$ to determine new values of x according to

$$x_{i+1} = -\tilde{G}_i^{-1} g_i \tag{36}$$

where $\tilde{G}_i$ depends upon the last n values of Δg and Δx. In this section we shall describe a recursive method whereby the gradient information available at each stage is utilised to adjust the offset vector computation, producing results analogous to eqn. (36).

Consider the first two steps in the offset vector method after which x_0, x_1, g_0 and g_1 are known. These 'points' are indicated for a two-dimensional case in **Figs. 3a** and **3b**. With Δx_1 and Δg_1 thereby determined, we can calculate the required first order change $\Delta x_1'$ in x to produce a desired change $\Delta g'$ in g in the direction of Δg_1:

$$\Delta x_1' = \frac{\Delta g'}{|\Delta g_1|}\,\Delta x_1. \tag{37}$$

Note that $\Delta g'$ is a scalar that may be either positive or negative. Out objective being to drive g to zero, to first order (approximately*), we can remove that component in the direction parallel to Δg_1 by defining

$$\left.\begin{aligned} \Delta g_1' &= -(g_1 . i_{\Delta g_1}) i_{\Delta g_1}; \quad i_{\Delta g_1} \equiv \frac{\Delta g_1}{|\Delta g_1|} \\ \Delta x_1' &= -(g_1 . i_{\Delta g_1}) \Delta x_1 / |\Delta g_1| \\ x_1' &= x_1 + \Delta x_1' \\ g_1' &= g_1 + \Delta g_1. \end{aligned}\right\} \tag{38}$$

These quantities are illustrated in Fig. 3. Note that $-(g_i . i_{\Delta g_i})$ in eqn. (38) plays the role of $\Delta g'$ in eqn. (37).

There is as yet no gradient information available in the direction normal to Δg_1 so, at this point, return to the offset vector algorithm. First, using eqn. (13) calculate the value γ_0' of the offset vector that corresponds to x_1':

$$\gamma_0' = \hat{g}(x_1'). \tag{39}$$

Assume that

$$g(x_1') \cong g_1 + \Delta g_1' \equiv g_1'; \tag{40}$$

note that exact equality does not hold because $\Delta g_1'$ is computed from a linearised analysis. Now let

$$\left.\begin{aligned} \gamma_1 &= \gamma_0' - g_1' \\ x_2 &= h(\gamma_1) \\ g_2 &= g(x_2). \end{aligned}\right\} \tag{41}$$

This completes a new step in the iteration process. Observe that the same number of evaluations of $g(x)$ are required as for the offset vector method. The difference is that x_2 is computed with the aid of an intermediate value x_1' that is calculated by a finite difference projection.

From x_2 and g_2 the quantities

$$\Delta x_2 = x_2 - x_1; \quad \Delta g_2 = g_2 - g_1 \tag{42}$$

are calculated as illustrated in Fig. 3. In the two-dimensional case Δx_1, Δx_2, Δg_1 and Δg_2 provide sufficient information to continue the search for x_∞ by a finite difference method alone, provided the Δx's and Δg's are independent. In higher dimensions we can proceed as before, calculating an intermediate x_2' based upon finite difference projections in both Δg_1 and Δg_2

* This is not an exact first order calculation because the gradient in the direction Δg_1 is computed from a finite difference.

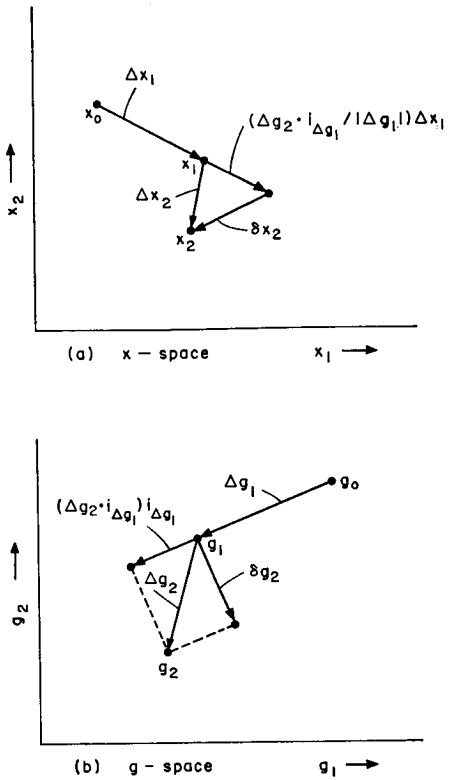


Fig. 4. Illustration of orthogonalisation of the vectors Δg_i with the associated transformation on the Δx_i

directions and using the offset vector to find corrections to x_2' in the remaining directions. Here we shall derive a recursion based upon orthogonalisation of the vectors Δg_1.

Suppose Δg_1, Δg_2, Δx_1 and Δx_2 are given as shown in **Fig. 4**. The component of Δg_2 orthogonal to Δg_1 is given by

$$\delta g_2 = \Delta g_2 - (\Delta g_2 . i_{\Delta g_1}) i_{\Delta g_1}.$$

According to eqn. (37), the associated change in x required to accomplish the increment δg_2 is given by

$$\delta x_2 = \Delta x_2 - (\Delta g_2 . i_{\Delta g_1} / |\Delta g_1|) \Delta x_1.$$

Defining $\delta x_1 = \Delta x_1$ and $\delta g_1 \equiv \Delta g_1$, we can calculate the change $\Delta g_2'$ required to drive the projection of an n-dimensional vector g_2 on the space of the orthogonal vectors δg_1 and δg_2 to zero. Requiring

$$(\Delta g_2' + g_2) . \delta g_i = 0; \quad i = 1, 2,$$

we have

$$\Delta g_2' = -(g_2 . i_{\delta g_1}) i_{\delta g_1} - (g_2 . i_{\delta g_2}) i_{\delta g_2}.$$

The associated change in x, $\Delta x_2'$, is given by

$$\Delta x_2' = -(g_2 . i_{\delta g_1}/|\delta g_1|)\delta x_1 - (g_2 . i_{\delta g_2}/|\delta g_2|)_{\delta x_2}.$$

Having $\Delta g_2'$ and $\Delta x_2'$, we can calculate x_1', g_2', γ_1, γ_2, x_3, and g_3 from eqns. (38), (39) and (41) by increasing the value of each subscript by one.

This reasoning leads to the following set of recursion relationships for deriving x_{i+1}, having $\{x_0, x_1, \ldots x_i\}$, $\{g_0, g_1, \ldots g_i\}$, orthogonal directions $\{\delta g_1, \delta g_2, \ldots \delta g_{i-1}\}$, and the corresponding set of 'influence' directions $\{\delta x_1, \delta x_2, \ldots \delta x_{i-1}\}$:

$$\left.\begin{aligned}
g_i &= g(x_i) \\
\Delta x_i &= x_i - x_{i-1} \\
\Delta g_i &= g_i - g_{i-1} \\
\delta g_i &= \Delta g_i - \sum_{j=1}^{i-1} (\Delta g_i . i_{\delta g_j}) i_{\delta g_j} \\
\delta x_i &= \Delta x_i - \sum_{j=1}^{i-1} (\Delta g_i . i_{\delta g_j}/|\delta g_j|)\Delta x_j \\
\Delta g_i' &= -\sum_{j=1}^{i} (g_i . i_{\delta g_j}) i_{\delta g_j} \\
\Delta x_i' &= -\sum_{j=1}^{i} (g_i . i_{\delta g_j}/|\delta g_j|)\delta x_j \\
x_i' &= x_i + \Delta x_i' \\
g_i' &= g_i + \Delta g_i' \\
\gamma_{i-1}' &= h(x_i') \\
\gamma_i &= \gamma_{i-1}' - g_i' \\
x_{i+1} &= h(\gamma_i).
\end{aligned}\right\} \quad (43)$$

To start the process, two iterations of the unmodified offset vector method are performed to provide values of x_0, x_1, g_0 and g_1. For $i \geq n$, we can discard all δx_j and δg_j for $j \leq i - n$; one set of directions is then effectively removed at each step to be replaced by δg_i and δx_i. Furthermore, for $i \geq n$ the last four expressions of eqn. (43) can be disregarded if we let

$$x_{i-1} \equiv x_i'; \quad i \geq n. \quad (44)$$

That is, a Newton–Raphson-like procedure, using approximate derivatives can be substituted for the offset vector method at the nth step.

8. Summary and conclusions

The offset vector method presented here is one that has been utilised to solve mathematical problems arising from special applications. The technique has evolved in this fashion because it requires knowledge of an approximate solution whose availability is dependent upon the physical situation. The purpose of this paper is to give the method more formal status as a numerical technique by presenting a recipe for its implementation, by developing criteria for convergence, and by illustrating its advantages through examples. It is found that the convergence rate is generally slower than that of second and higher order methods, but each iteration is relatively rapid to perform. Possible applications are those where few iterations are required or as a starting procedure for higher order methods when many iterations are necessary.

References

BATTIN, R. H. (1964a). 'Explicit and Unified Methods of Spacecraft Guidance Applied to a Lunar Mission' (Warsaw, Poland: XVth International Astronautical Congress).

BATTIN, R. H. (1964b). *Astronautical Guidance*, McGraw-Hill, Inc., New York.

GODAL, T. (1961). 'Method for Determining the Initial Velocity Corresponding to a Given Time of Free Flight Transfer between Given Points in a Simple Gravitational Field', *Astronautik*, Vol. 2, p. 183.

PRICE, C. F., and BOYLAN, ROBERT F. (1964). 'A Method for Computing the Correlated Velocity Vector for a Body in an Oblate Earth Gravitational Field', *Instrumentation Laboratory Report* E-1645, Cambridge, Mass.

PRICE, CHARLES F. (1967). 'An Indirect Numerical Method Based on Offset Vectors for Obtaining Solutions to Optimal Control Problems', XVIII IAF Congress Proceedings, Belgrade, Yugoslavia, Sept. 1967. See also the same title, Experimental Astronomy Report RE-30, Sept. 1967, Massachusetts Institute of Technology, Cambridge, Massachusetts, U.S.A.

SLATER, GARY L., and STERN, ROBERT G. (1966). 'Simplified Midcourse Guidance Techniques', *Experimental Astronomy Laboratory Report* RE-18, Cambridge, Mass.

TODD, JOHN, editor (1962). *Survey of Numerical Analysis*, McGraw-Hill Book Co., Inc., Chapter 7.

MINIMUM IMPULSE GUIDANCE

By T. N. Edelbaum
Analytical Mechanics Associates, Inc.
Cambridge, Massachusetts

MINIMUM IMPULSE GUIDANCE*

T. N. Edelbaum
Analytical Mechanics Associates, Inc.
Cambridge, Massachusetts

Abstract

A linearized theory is developed for minimum fuel guidance in the neighborhood of a minimum-fuel space trajectory. The thrust magnitude is unrestricted so that the thrust is applied impulsively on both the nominal trajectory and the neighboring optimal trajectories. The analysis allows for additional small midcourse impulses as well as for small changes in the magnitude, direction, and timing of the nominal impulses. The fuel is minimized by determining the trajectory which requires the minimum total velocity change when summed over all the impulses.

The analysis is deterministic and applies to arbitrary time-varying gravitational fields. Three separate time-open problems are treated; rendezvous, orbit transfer, and orbit transfer with tangential nominal impulses.

*Performed under contract NAS 12-114, presented at AIAA 7th Aerospace Sciences Meeting, January 1969, Preprint No. 69-74.

List of Symbols

$\delta\bar{R}$	Position deviation
t_f	Nominal final time
δt	Time deviation
u_c	Component of midcourse impulse in the critical plane
u_{nc}	Component of midcourse impulse in the noncritical direction
$\bar{V}_T$	Velocity of target trajectory
$\bar{V}_I$	Velocity of nominal interception trajectory
$\overline{\Delta V}$	Nominal terminal impulsive velocity change
$\Sigma\delta V$	Total change in impulsive velocity cost

MINIMUM IMPULSE GUIDANCE

Introduction

This is the second of a series of papers on minimum fuel guidance of high-thrust rockets. The first paper (Ref. 1) illustrated the general approach by treating the particular problem of guidance from a hyperbolic to a circular orbit. The succeeding papers are intended to generalize this approach to more general classes of guidance problems. This generalization will be carried out in several stages. The present paper will consider the general case of time-open impulsive guidance. Later papers will extend the analysis to finite thrust.

There is a well-developed theory for minimum fuel impulsive guidance, e.g., Refs. 2, 3 and 4. However, these references consider only the case of an unpowered nominal trajectory. The nominal trajectory around which the analysis is linearized is a coasting arc. The present paper is intended to generalize these results to nominal trajectories containing one or more finite impulses. The analysis will consider three different problems. The first problem to be treated will be minimum fuel guidance for time-open rendezvous. The second problem will be time-open orbit transfer, and the third problem will be an important special case of the second, where one or more of the finite impulses is tangent to the velocity vector.

Mathematical Model

The analysis of the present paper is linearized about a nominal trajectory containing one, or more, finite impulsive velocity changes. This nominal trajectory must be an optimal trajectory minimizing the sum of the absolute magnitude of its impulses for transfer between its terminal states. The problem considered is the deterministic problem of determining the minimum impulse transfer from a given state in a close neighborhood of the nominal state at a given initial time to the terminal state with time open. The nominal trajectory may lie in a general time-varying gravitational field. The analysis is a first order analysis neglecting second order terms. It is analogous to the neighboring optimal guidance schemes developed for smooth optimization problems without corners. The problem is complicated by the possession of corners and the possibility of introducing additional impulses. However, the problem is simplified because it is a first order analysis. In general, the problem will be to guide the vehicle from a given initial state at a given initial time to a final time in the near vicinity of the nominal terminal time. For the orbit transfer problem the final time may be allowed to become arbitrarily large; and it may also be possible to extend the initial time arbitrarily far backwards in time.

MINIMUM IMPULSE GUIDANCE

Analysis

I. Time-Open Rendezvous

The key concept in analyzing minimum-impulse guidance for time-open rendezvous is the concept of a noncritical direction. This concept was originally developed for use in interception problems rather than rendezvous (Refs. 2 and 5) but is also useful in analyzing rendezvous. Consider the case where the nominal trajectory has a single finite impulse which accomplishes rendezvous at a nominal terminal time. If rendezvous were to be accomplished at a slightly earlier time δt, then the point at which rendezvous is accomplished must be displaced by the negative product of the target velocity vector and the time change.

$$\delta \bar{R} = - \bar{V}_T \delta t \qquad @\ t = t_f - \delta t \tag{1}$$

This position is reached by the intercepter at an earlier time than the nominal arrival time. If the trajectory of the intercepter were continued to the nominal arrival time, it would have the position given by Eq. (2) and shown on Fig. 1.

$$\delta \bar{R}_I = - \bar{V}_T \delta t + \bar{V}_I \delta t = - \overline{\Delta V}\, \delta t \qquad @\ t = t_f \tag{2}$$

This indicates that, if the intercepter will intercept a specified line in space at the nominal arrival time, then it will (to first order) also intercept the target at a somewhat earlier or later time. This specified line passes through the nominal arrival point and has the direction of the nominal finite impulse. This direction through the nominal arrival point is known as the noncritical direction

at the nominal arrival time. It represents the one permissible direction of position variation which will still lead to rendezvous. This noncritical direction may also be propagated backward in time by use of the state transition matrix. It will then define a noncritical direction at any point along the nominal trajectory.

In order to effect rendezvous, it is necessary to control the two components of position variation in the plane normal to the noncritical direction. This plane is known as the critical plane. Once the terminal position of the target vehicle and the rendezvous vehicle has been matched by reducing the position deviations in the critical plane to zero, rendezvous is accomplished by a finite impulse which nulls the difference between the target and intercepter velocities. To first order, only one component of terminal impulse variation adds linearly to the cost; that in the direction of the nominal impulse. Any small deviations in the velocity vector normal to this direction may be cancelled by small rotations of the nominal terminal impulse. Such rotations only increase cost to second order and may be neglected in a first order analysis.

The foregoing considerations indicate that only two components of position and one component of velocity at the nominal final time must be controlled for time-open rendezvous. This reduces the original 6-dimensional parameter space to a 3-dimensional parameter space. If there is only one

finite impulse, then the analysis for unpowered nominals in Refs. 2, 3 and 4 may be applied without change to this 3-dimensional parameter space. That analysis indicates that the optimum solution has no more than three impulses. One of these impulses will represent a variation in the magnitude of the nominal impulse so that there are, at most, two midcourse impulses.

The required position correction at the nominal terminal time may be accomplished with a single midcourse impulse. If this corrective impulse occurs at a specified time, then the optimum direction of this impulse may easily be calculated. One component of the impulse will produce the position correction. This component will lie in the critical plane. There will also be a component of the midcourse impulse in the noncritical direction. This component will be used to reduce the magnitude of the large terminal impulse and will result in an overall saving in impulse magnitude and fuel. The total change in impulsive velocity is given by Eq. (3).

$$\Sigma \delta V = \sqrt{u_c^2 + u_{nc}^2} - \frac{\partial |\overline{\Delta V}|}{\partial u_{nc}} u_{nc} - \frac{\partial |\overline{\Delta V}|}{\partial u_c} u_c \tag{3}$$

The optimum magnitude of the velocity component in the noncritical direction may be found by differentiating Eq. (3), and solving for the stationary minimum point given by Eq. (4).

$$u_{nc}^* = \frac{\frac{\partial |\overline{\Delta V}|}{\partial u_{nc}} |u_c|}{\sqrt{1 - \left[\frac{\partial |\overline{\Delta V}|}{\partial u_{nc}}\right]^2}} \tag{4}$$

The total cost of the optimum correction at a specified time is given by Eq. (5).

$$\Sigma\delta V^* = \sqrt{1 - \left[\frac{\partial|\Delta V|}{\partial u_{nc}}\right]^2}\,|u_c| - \frac{\partial|\Delta V|}{\partial u_c}\,u_c \tag{5}$$

In the particular case treated in Ref. 1, the midcourse correction should be made as early as possible and there will be only one midcourse impulse for the minimum fuel solution. This behavior will be typical of most cases as the time approaches the terminal time. However, in other cases as many as two midcourse impulses will be required to minimize the fuel consumption. It is also possible that a single impulse at a time later than the time under consideration may be optimum. There are both direct and indirect approaches to this optimization problem. The indirect method calculates the primer vector (Refs. 6 and 7) from the direction given by the optimum direction of a single midcourse impulse at the current time to the terminal impulse at the terminal time. If this vector is less than unity at all intermediate points, then the single correction will be the absolute minimum fuel solution.

The direct method is a constructive approach utilizing the convex hull of the reachable set of terminal states (Ref. 2). This reachable set is constructed in a parameter space defined by the change in the terminal impulse magnitude and by the two position components in the terminal critical plane. Each of these parameters is normalized by the magnitude of the midcourse velocity change. An optimum maneuver must lie on the convex hull of the

reachable sets in this space. The set of all impulse directions at a given time will define an ellipsoid in the parameter space. Equations (4) and (5) will define a generator of a cone which is tangent to the ellipsoid and whose apex is at minus one on the velocity axis (see Fig. 2). If a single correction at the earliest possible time is optimal, then the cones for all subsequent times will lie inside the initial cone. If two midcourse corrections are required, then the convex hull of all the cones will have a plane as one of its bounding surfaces. If a single correction at a later time is optimal, then one of the later cones will project through the cone corresponding to the initial time. The geometric construction for these cases may be reduced to a 2-dimensional construction by using the traces of the cone on the plane of the position variations. In exceptional cases where such traces do not produce closed figures, it may be necessary to use another plane that passes through the cones.

If the nominal trajectory contains one or more large impulses before the final impulse, then all necessary corrections may be made by utilizing small variations in these impulses. It is only necessary to consider small variations of timing and direction of these impulses. Such variations allow control of one component of position and two components of velocity at the time of the impulse. These three components may then be propagated to the terminal state by means of the state transition matrix. Except in exceptional cases it will be possible to control all three required components of the terminal state by this means. This control will (to first order) produce no increase in cost. This is shown by the

fact that the primer vector passing through the two impulses of the optimal nominal trajectory is stationary with respect to small variations in impulse timing and direction.

II. Time-Open Orbit Transfer

If the object of the mission is orbit transfer rather than rendezvous, the particular phasing of the vehicle in the final orbit is unspecified. This means that there will be a set of noncritical directions arising from all points on the target orbit in the vicinity of the nominal terminal time. This set of directions will to first order define a plane in which will lie the velocity vectors of both the target orbit and the transfer orbit at the nominal terminal time. All trajectories which are close neighbors of the nominal trajectory and which touch this noncritical plane at the nominal terminal time will also intersect the target trajectory at a time close to the nominal terminal time. For the orbit transfer problem it is only necessary to control the one component of terminal position in the critical direction which is normal to the noncritical plane. The parameter space which must be considered is only 2-dimensional, containing one position component and one velocity component. There will be at most one midcourse impulse in addition to small variations in the terminal impulse. The optimum midcourse impulse may occur at a time other than the earliest possible time. In fact, in some cases this single midcourse impulse should occur in the neighborhood of the terminal orbit rather than in the neighborhood of the transfer

orbit and at a time later than the time of the nominal terminal impulse. The latter case is easily analyzed by considering the set of reachable states in the vicinity of the terminal orbit, as well as in the vicinity of the transfer orbit.

III. Time-Open Orbit Transfer with Tangential Impulses

In many orbit transfer problems, such as the well-known Hohmann transfer, the impulses are applied tangent to the velocity vector. In such a case the noncritical plane of the preceding section becomes undefined and it is once again necessary to consider a 3-dimensional parameter space possessing two components of position variation. This case is similar to the case of time-open rendezvous and possesses a noncritical direction and a critical plane. As in the preceding section, it may be desirable to consider midcourse impulses in the terminal orbit as well as in the transfer orbit. It is possible to have a midcourse impulse before the major transfer impulse in the neighborhood of the transfer orbit, as well as a post-terminal-time midcourse impulse in the neighborhood of the nominal terminal orbit. If there are one or more large impulses on the nominal trajectory before the terminal impulse, then variations in the timing and direction of these impulses may be used to control the trajectory. In the particular case of a Hohmann transfer, these variations will not be sufficient to control all out-of-plane deviations because the two impulses are located at singularities of the state transition matrix. In this case it will be necessary to utilize midcourse impulses in either the transfer orbit or one of the terminal orbits for controlling the out-of-plane component of the terminal position variation.

Conclusions

(1) Minimum impulse time-open rendezvous in the neighborhood of an optimal nominal trajectory requires at most two small midcourse impulses if the nominal trajectory possesses one large finite impulse. Two midcourse impulses may be required if either the nominal trajectory or the deviations from it are nonplanar. If both the trajectory and deviations are planar, not more than one midcourse impulse will be required to realize minimum total impulse.

(2) Minimum fuel, time-open orbit transfer in the near vicinity of an optimum nominal requires at most one small midcourse impulse if the nominal trajectory contains at least one finite impulse which is not tangent to the velocity vector. If both the nominal trajectory and the small deviations from it lie in the same plane, there will be no small midcourse impulse. In the latter case, the first order minimum fuel solution will be a single impulse at the intersection of the two orbits.

(3) For both time-open rendezvous and orbit transfer with two or more finite impulses, no midcourse impulse will be required unless the finite impulses occur at singularities of the state transition matrix.

References

1. Edelbaum, T. N., "Optimal Guidance from Hyperbolic to Circular Orbits", Second Compilation of Papers on Trajectory Analysis and Guidance Theory, NASA PM-67-21, 1968, pp. 27-50; also to appear in the Proceedings of the Colloquium on Advanced Problems and Methods for Space Flight Optimization, Liege, Belgium, June 19-23, 1967.

2. Potter, J. E. and Stern, R. E., "Optimization of Midcourse Velocity Corrections", Proc. of the IFAC Symposium on Automatic Control in Peaceful Uses of Space, Stavanger, Norway, J. A. Aseltine, ed., Plenum Press, New York, June 1965.

3. Neustadt, L. W., "Optimization, A Moment Problem and Nonlinear Programming", Society for Industrial and Applied Mathematics (SIAM), Journal on Control, Vol. 2, No. 1, 1964, pp. 33-89.

4. Platonov, A. K., "Investigation of the Properties of Correction Maneuvers in Interplanetary Flights", Cosmic Research, 4, 1966, pp. 587-607.

5. Noton, A. R. M., Cutting, E., and Barnes, F. L., "Analysis of Radio-Command Midcourse Guidance", JPL Technical Report No. 32-28, Sept. 1960.

6. Lion, P. M., "A Primer on the Primer", Second Compilation of Papers on Trajectory Analysis and Guidance Tneory, NASA PM-67-21, 1968, pp. 85-105.

7. Lawden, D. F., Optimal Trajectories for Space Navigation, Butterworths, London, 1963.

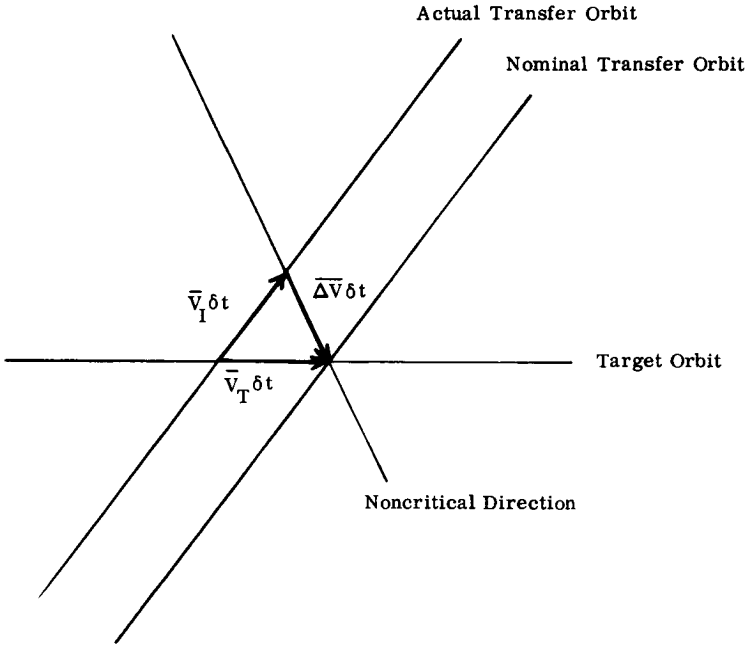


Fig. 1 Intercept Geometry

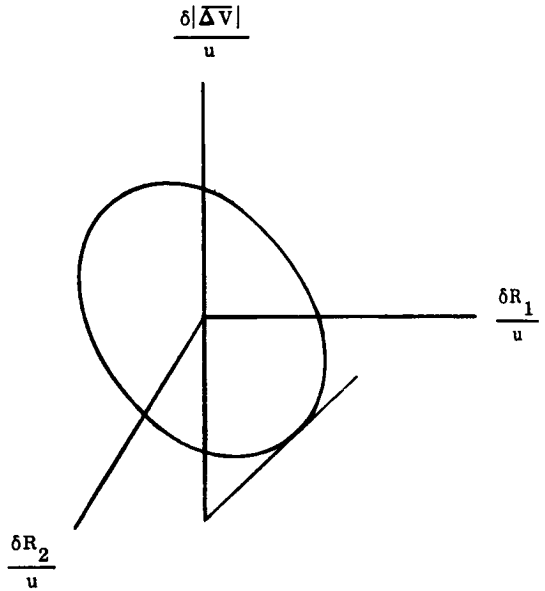


Fig. 2 Parameter Space

IMPROVEMENT OF THE SPHEROIDAL METHOD FOR ARTIFICIAL SATELLITES

By John P. Vinti
Massachusetts Institute of Technology
Cambridge, Massachusetts

Improvement of the Spheroidal Method for Artificial Satellites

John P. Vinti

Experimental Astronomy Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

ABSTRACT

Objections to applying the spheroidal method to calculate a polar orbit of an artificial satellite are easily overcome.

Previous papers have already treated the behavior in an exactly polar orbit of the right ascension ϕ, the coordinate for which the difficulty supposedly occurs. Just as in the Keplerian problem, it remains constant, except for jumps of 180° at a pole.

There remains the case of an almost polar orbit, for which the calculation of ϕ may be inaccurate near a pole, unless one takes special precautions. The present paper first simplifies the expression for ϕ for all orbits, polar or not, and then shows how to avoid the difficulty altogether, by solving directly for rectangular coordinates and velocities. These considerations apply both to papers by the author and by Izsak on the original spheroidal method and to the author's later papers incorporating the third zonal harmonic into the spheroidal potential.

The present paper simplifies orbital calculations by the spheroidal method for satellite orbits with all inclinations. Its main points are the bypassing of the right ascension and the avoidance of differences of almost equal quantities, so that all calculations become well-conditioned.

IMPROVEMENT OF THE SPHEROIDAL METHOD FOR ARTIFICIAL SATELLITES

1. INTRODUCTION

Objections have sometimes been made to applying the author's spheroidal method to calculate a polar orbit of an artificial satellite. The coordinates that appear are ρ, for which the level surfaces are oblate spheroids, η, for which they are hyperboloids of one sheet, and the right ascension ϕ. The apparent difficulty in a polar orbit arises only in ϕ and then only at a pole.

For an exactly polar orbit I have already shown by limiting processes in V1961a and V1961b[1] that the spheroidal potential leads to ϕ = constant, except at a pole, where it jumps by $\pm$ 180°, accordingly as we call the orbit direct or retrograde, respectively. This is the expected behavior, just the same as for a Keplerian orbit, so that no real difficulty appears. It holds whether or not the model takes into account the third zonal harmonic, with coefficient J_3.

Although the difficulty was easily disposed of, without tedious numerical calculations, for an exactly polar orbit, one might still claim that it remains troublesome for an almost polar orbit. For such an orbit the calculation of ϕ involves a small denominator which almost vanishes near a pole. One then may very likely lose accuracy in passing by the pole or have to use special procedures which will increase computer time and storage demands and which will not elsewhere be necessary. The present paper shows how to avoid such difficulties.

2. THE AUTHOR'S SPHEROIDAL SOLUTION, WITHOUT J_3

The notation in this section is that of V1961a, corrections of which are to be found in Walden and Watson 1967, p. 16. The rectangular coordinates X, Y, Z satisfy

$$X + iY = (\rho^2+c^2)^{1/2} (1-\eta^2)^{1/2} \exp i\phi \qquad (1.1)$$

$$Z = \rho\eta \qquad (1.2)$$

Now by (8.50) of V1961a,

$$\phi = \Omega' + F , \qquad (2)$$

1. The initial V refers to the author's own papers.

where F is that part of the expression which varies rapidly near a pole. Here Ω' is given by Eq. (9) of the present paper and

$$F = K\chi \tag{3.1}$$

$$K = |K| \operatorname{sgn} \alpha_3 , \tag{3.2}$$

where

$$K^2 = \alpha_3^2\eta_0^2\eta_2^2(\alpha_2^2-\alpha_3^2)^{-1}(\eta_0^2+\eta_2^2-1-\eta_0^2\eta_2^2)^{-1} \tag{4}$$

But

$$\eta_0^2+\eta_2^2 = 1+\alpha_2^2(-2\alpha_1 c^2)^{-1} \qquad \text{(4.1 of V 1961a)}$$

$$\eta_0^2\eta_2^2 = (\alpha_2^2-\alpha_3^2)(-2\alpha_1 c^2)^{-1} \qquad \text{(4.2 of V 1961a)}$$

It follows that $K^2 = 1$, so that

$$K = \operatorname{sgn} \alpha_3 = \pm 1 \tag{5}$$

for direct or retrograde orbits, respectively, in order that the right ascension ϕ may correspondingly either always increase or always decrease. Then

$$\phi = \Omega' + \chi \operatorname{sgn} \alpha_3 \tag{6}$$

is an exact equation for all orbits, with the spheroidal model. This is in contradistinction to the results of V1961a, where it was only shown to hold for polar orbits. Thus the present work simplifies all calculations with the spheroidal model.

To find the rectangular coordinates X and Y directly, without first calculating ϕ, insert (6) into (1.1), use

$$\exp i\chi = (1-\eta_0^2 \sin^2 \psi)^{-\frac{1}{2}} (\cos \psi + i\sqrt{1-\eta_0^2} \sin \psi) \tag{7}$$

from the last paragraph of V1961b, and then put $\eta=\eta_0 \sin \psi$ and $(1-\eta_0^2)^{1/2} = |\cos I|$, from (6.4) and (4.7) of V1961a. The troublesome denominator $(1-\eta^2)^{1/2}$ then cancels out, with the result

$$X+iY = (\rho^2+c^2)^{\frac{1}{2}} (\cos\psi + i \cos I \sin\psi)\exp i\Omega' \tag{8}$$

for all orbits, direct or retrograde. Here

$$\Omega' = \beta_3+\alpha_3(\alpha_2^2-\alpha_3^2)^{-\frac{1}{2}} \eta_0(B_3\psi+ \tfrac{3}{32} \eta_0^2\eta_2^{-4} \sin 2\psi)$$
$$-c^2\alpha_3(-2\alpha_1)^{-1/2}(A_3\nu+ \sum_{k=1}^{4} A_{3k}\sin k\nu) \tag{9}$$

Separately

$$X=(\rho^2+c^2)^{\frac{1}{2}} (\cos\Omega' \cos\psi-\sin\Omega' \cos I \sin\psi) \tag{10.1}$$

$$Y=(\rho^2+c^2)^{\frac{1}{2}} (\sin\Omega' \cos\psi + \cos\Omega' \cos I \sin\psi) \tag{10.2}$$

These expression contain no singularities or rapidly varying quantities, so that there is thus never any difficulty with a polar or almost polar orbit. For a strictly polar orbit cos I and α_3 both vanish, so that $\Omega'=\beta_3$ and

$$X+iY = (\rho^2+c^2)^{1/2} \cos\psi \exp i\beta_3 \tag{11}$$

3. Izsak's Spheroidal Solution

Although Izsak (1960, 1963) suggested using a slowly rotating reference plane to avoid the polar difficulty, actually the same transformations hold for his solution of the spheroidal problem. For the sake of accessibility, I shall refer to his 1963 paper. In making the comparison, note that my symbols are to be changed as follows: $\phi \to \alpha$, $\eta \to \sigma$, $\eta_0 \to s$, and $\beta_3 \to \Omega_*$; others remain the same. Then, with use of Izsak's Eqs. (3), (91), (37), and (63), one finds again the equivalent of the present Eqs. (10) for the rectangular coordinates X and Y. Note that Izsak's expression for Ω' contains $(1-s^2)^{1/2}$ in the numerator and $1-e^2$ in the denominator of each term except Ω_*. The $1-e^2$ in such a denominator does not necessarily produce a singularity as $e \to 1$, since each $(1-e^2)^{-1}$ is multiplied by $\nu=c/a$ and $p = a(1-e^2)$ is a quantity analogous to the semi-latus rectum in a Keplerian orbit. In such an orbit

$p > 0$ for any orbit that does not intersect the center of the planet, even if e=1. Incidentally, the same powers of p occur in the coefficients B_3, A_3 and the A_{3k}'s of V1961a.

4. Isolation of the Right Ascension

In either solution, the quantity here called χ is the sensitive part of the expression for the right ascension ϕ. If one actually wants values of ϕ near a pole in an almost polar orbit, it is better to rewrite Eq. (7) as

$$\exp i\chi = (\cos^2\psi+\cos^2 I \sin^2\psi)^{-\frac{1}{2}} (\cos\psi +i|\cos I|\sin \psi) \tag{12}$$

One thus avoids calculating the difference of two almost equal numbers in the denominator. Then ϕ is given by (6) and (12).

5. Velocity Components, with $J_3=0$

On taking the logarithmic derivative of (8) and multiplying the result by X+iY, we find

$$\dot{X}+i\dot{Y} = \left(\frac{\rho\dot{\rho}}{\rho^2+c^2} + i\dot{\Omega}' \right)(X+iY)+(\rho^2+c^2)^{\frac{1}{2}} (-\sin\psi +i \cos I \cos \psi)\dot{\psi}\epsilon^{i\Omega'} \tag{13}$$

so that

$$\dot{X} = \frac{\rho\dot{\rho}}{\rho^2+c^2} X-Y\dot{\Omega}'+(\rho^2+c^2)^{\frac{1}{2}} (-\sin \psi \cos \Omega'-\cos I \cos \psi \sin \Omega')\dot{\psi} \tag{14.1}$$

$$\dot{Y} = \frac{\rho\dot{\rho}}{\rho^2+c^2} Y+\chi\dot{\Omega}'+(\rho^2+c^2)^{\frac{1}{2}} (-\sin \psi \sin \Omega' + \cos I \cos \psi \cos \Omega')\dot{\psi} \tag{14.2}$$

Differentiation of (1.2) gives

$$\dot{z} = \eta\dot{\rho}+\rho\dot{\eta} = \eta\dot{\rho}+\eta_0\rho \cos \psi \dot{\psi} \tag{15}$$

These equations contain neither small denominators nor differences of almost equal quantities. Here

$$\dot{\rho} = ae\left(\frac{\mu}{a_0}\right)^{\frac{1}{2}} (\rho^2+A\rho+B)^{\frac{1}{2}} (\rho^2+c^2\eta^2)^{-1} \sin E \tag{16}$$

from p. 6 of Bonavito 1962, and

$$\dot{\eta} = \eta_0 \cos\psi \quad \dot{\psi} = (\alpha_2^2-\alpha_3^2)^{\frac{1}{2}} (1-q^2\sin^2\psi)^{\frac{1}{2}} (\rho^2+c^2\eta^2)^{-1} \cos\psi, \tag{17}$$

from p. 15 of Walden 1967, after a few transformations. Here $q=\eta_0/\eta_2$. Then

$$\dot{\psi} = \eta_0^{-1}(\alpha_2^2-\alpha_3^2)^{\frac{1}{2}} (\rho^2+c^2\eta^2)^{-1} (1-q^2\sin^2\psi)^{\frac{1}{2}} \tag{18}$$

Finally, by Eq. (9) of the present paper,

$$\dot{\Omega}' = \alpha_3(\alpha_2^2-\alpha_3^2)^{-\frac{1}{2}} \eta_0 (B_3+ \tfrac{3}{16} \eta_0^2\eta_2^{-4}\cos 2\psi)\dot{\psi}$$

$$-c^2\alpha_3(-2\alpha_1)^{-\frac{1}{2}} (A_3+ \sum_{k=1}^{4} kA_{3k}\cos kv)\dot{v} \tag{19}$$

Thus we also need $\dot{v}$. With

$$\rho = (1+e \cos v)^{-1}p, \tag{20}$$

from (5.12) of V1961a, where $p=a(1-e^2)$, we find

$$\dot{\rho} = \frac{e}{p} \rho^2 \sin v \, \dot{v} \tag{21}$$

Comparison of (16) and (21), with use of the anomaly connection

$$\sin E = \frac{\rho}{p} (1-e^2)^{\frac{1}{2}} \sin v \tag{22}$$

then gives

$$\dot{v} = \frac{a}{\rho}\left[\frac{\mu(1-e^2)}{a_0}\right]^{\frac{1}{2}} \frac{(\rho^2+A\rho+B)^{\frac{1}{2}}}{\rho^2+c^2\eta^2} \tag{23}$$

Eqs. (14), (15), (16), (18), (19), and (23) then give the complete

algorithm for finding the velocity components in the spheroidal model, when J_3 is not included.

6. The Author's Spheroidal Solution, with J_3

The notation in this section is that of V1966, corrections of which are to be found in Walden and Watson 1967, pp. 19, 20, 22, 27, and 31. With this solution

$$\phi = \Omega' + G \operatorname{sgn} \alpha_3, \tag{24}$$

where Ω' is given in Eq. (41.4) of the present paper and where G is given by Eq. (150) of V1966, viz

$$G = |\alpha_3| \alpha_2^{-1} u^{\frac{1}{2}} (1-S)^{-\frac{1}{2}} [(h_1+h_2)\chi_0 + (h_1-h_2)\chi_1] \tag{25}$$

From Eq.(158) of V1966, we have

$$(h_1+h_2)\chi_0 + (h_1-h_2)\chi_1 = 2^{-1}(1-C_2)^{\frac{1}{2}} [(1-C_2)^2 - C_1^2]^{-\frac{1}{2}} (E_2'+E_3') \tag{26}$$

If u is a solution of the cubic equation (27) of V1966, then by (32.1) and (32.2) of that paper

$$C_2 = \frac{c^2 u}{a_0 p_0} \ (16), \quad C_1 = 2u\delta p_0^{-1}(1-C_2 S)^{-1}(1-C_2), \tag{27}$$

so that

$$\frac{(1-C_2)^2 - C_1^2}{1-C_2} = \frac{u}{1-S}\left[\left(\frac{1}{u} - \frac{c^2}{a_0 p_0}\right)(1-S) - R\right], \tag{28.1}$$

where

$$R \equiv \left(\frac{1}{u} - \frac{c^2}{a_0 p_0} S\right)^{-2} \left(\frac{2\delta}{p_0}\right)^2 (1-S)\left(\frac{1}{u} - \frac{c^2}{a_0 p_0}\right) \tag{28.2}$$

By (27) of V1966, however,

$$R = \frac{1}{u} - 1 - \frac{c^2}{a_0 p_0}(1-S) \tag{29}$$

Insertion of (29) into (28) then shows that

$$(1-C_2)^{-1}[(1-C_2)^2-C_1^2] = (1-S)^{-1}(u-S) , \tag{30}$$

which, with (26), gives

$$(h_1+h_2)\chi_0+(h_1-h_2)\chi_1 = 2^{-1}(u-S)^{-\frac{1}{2}}(1-S)^{\frac{1}{2}}(E_2'+E_3') \tag{31}$$

Now, by Eqs. (21.2), (18), and (26) of V1966, for all orbits, direct or retrograde,

$$|\alpha_3|\alpha_2^{-1}u^{\frac{1}{2}} = (u-S)^{\frac{1}{2}} \tag{32}$$

Then, from (25), (31), and (32),

$$G = \tfrac{1}{2}(E_2'+E_3') \tag{33}$$

for all orbits, polar or not, and direct or retrograde. This is the same as the expression given in Eqs.(159) of V1966 for the sensitive part of ϕ in the case of a polar orbit. Here, however, we have shown that it holds for all orbits.

To evaluate G, place $E_2' = E_2'(\psi+\pi/2)$ and $E_3' = E_3'(\psi-\frac{\pi}{2})$ into Eqs.(104) of V1966. The results are

$$\cos E_2' = \frac{e_2-\sin\psi}{1-e_2\sin\psi} \qquad \cos E_3' = \frac{e_3+\sin\psi}{1+e_3\sin\psi}$$

$$\sin E_2' = \frac{(1-e_2^2)^{\frac{1}{2}}\cos\psi}{1-e_2\sin\psi} \qquad \sin E_3' = -\frac{(1-e_3^2)^{\frac{1}{2}}\cos\psi}{1+e_3\sin\psi} , \tag{34}$$

where

$$e_2=(1-P)^{-1}Q, \quad e_3=(1+P)^{-1}Q , \quad Q^2=P^2+S , \tag{35}$$

with $0 \leq e_3 \leq e_2 \leq 1$, by Eqs. (100) and (47) of V1966. Then, by (33),

$$\cos(E_2'+E_3') = \cos 2G = 2\cos^2 G-1 \tag{36}$$

From (34) and (36) it then follows that

$$\cos G = k(\psi)\,\frac{-\frac{1}{2}^{\frac{1}{2}}\left(1+e_2e_3+\sqrt{(1-e_2^2)(1-e_3^2)}\right)^{\frac{1}{2}}\cos\psi}{[(1-e_2\sin\psi)(1+e_3\sin\psi)]^{\frac{1}{2}}}\,, \qquad (37)$$

where $k(\psi)=\pm 1$.

We now show that $k(\psi)=1$ for all ψ. First note that $E_2'(y)$ is related to y in the same way that an eccentric anomaly is related to a true anomaly. The same holds for $E_3'(y)$. Thus each increases as y increases, by Eq. (160) of V1966, so that $G\equiv 2^{-1}\times[E_2'(\psi+\pi/2)+E_3'(\psi-\pi/2)]$ is a continuous monotonically increasing function of ψ.

Also, from the definitions, $E_2'(y)$ and $E_3'(y)$ are both equal to $n\pi$ for $y = n\pi$. Thus

$$G = \psi \quad \text{for } \psi = (n+\tfrac{1}{2})\pi, \quad (n=0,1,2\ldots) \qquad (38)$$

so that cos ψ and cos G both vanish for $\psi=(n+\frac{1}{2})\pi$. Now consider a small interval $(n+\frac{1}{2})\pi-\epsilon \leq \psi \leq (n+\frac{1}{2})\pi+\epsilon$. Since G always increases with increase in ψ, the corresponding changes $\Delta\cos\psi$ and $\Delta\cos G$ are both negative if n is even and both positive if n is odd. Thus $k(\psi) > 0$ over any such inteval. But $k(\psi) = \pm 1$ for all ψ and since cos G and thus $k(\psi)$ are continuous functions of ψ, it follows that

$$k(\psi) = 1 \text{ for all } \psi \qquad (39)$$

Before we rewrite (37) with omission of $k(\psi)$, let us first simplify it. To do so, note that by (35) and by (48) of V1966, which is

$$\eta = P + Q\sin\psi, \qquad (40)$$

we obtain

$$(1-e_2\sin\psi)(1+e_3\sin\psi) = (1-P^2)^{-1}(1-\eta^2) \qquad (41)$$

Now from Eq. (32.3) of V1966

$$2P = r(1-S)\delta, \qquad (42.1)$$

where

$$\delta = \frac{r_e}{2} J_2^{-1} |J_3| \tag{42.2}$$

$$r = 2(1-C_2S)^{-1}u/p_o \tag{42.3}$$

Thus $\delta = O(J_2)$ and $r = O(1)$. Eqs. (35) and (42) then show that

$$1+e_2e_3 + \sqrt{(1-e_2^2)(1-e_3^2)} = (1-P^2)^{-1}(1+S+(1-S)\sqrt{1-r^2\delta^2}) \tag{43}$$

On inserting (39), (41), and (43) into (37), we find

$$\cos G = 2^{-\frac{1}{2}}(1-\eta^2)^{-\frac{1}{2}}[1+S+(1-S)\sqrt{1-r^2\delta^2}]^{\frac{1}{2}}\cos\psi \tag{44}$$

We also need sin G in calculating rectangular coordinates. To evaluate it unambiguously first note that

$$2 \sin G \cos G = \sin (E_2'+E_3') \tag{45}$$

$$=(1-\eta^2)^{-1}(1-P^2)[(e_3\sqrt{1-e_2^2} - e_2\sqrt{1-e_3^2})+(\sqrt{1-e_2^2}+\sqrt{1-e_3^2})\sin\psi]\cos\psi \tag{46}$$

by (24) and (31). Then from (35), (42), (44), (45), and (46) it follows that

$$\sin G = \frac{2^{-\frac{1}{2}}(1-S)^{\frac{1}{2}}}{(1-\eta^2)^{\frac{1}{2}}} \frac{\{Q(\sqrt{1-r\delta} - \sqrt{1+r\delta})+[(1+P)\sqrt{1-r\delta} + (1-P)\sqrt{1+r\delta}]\sin\psi\}}{[1+S+(1-S)\sqrt{1-r^2\delta^2}]^{\frac{1}{2}}} \tag{47}$$

To check this, note that for $J_3 = 0$ we have $\delta = 0$, $P = 0$, $Q = S^{1/2}$, and $S = \sin^2 I$, so that (47) then reduces to

$$\sin G = (1-\eta^2)^{-\frac{1}{2}}|\cos I|\sin\psi, \tag{48}$$

agreeing with (7) for sin χ.

If one really wants values of the right ascension near a pole, one can use (24), (44), and (47).

It is then advisable, however, to rewrite the $1-\eta^2$ in the denominator by using (35) and (40). One finds

$$1-\eta^2 = \cos^2\psi-(P^2+2PQ\sin\psi)+(1-S-P^2)\sin^2\psi, \tag{49}$$

resulting in the same kind of simplification near a pole as does (12).

Near a pole in a nearly polar orbit the term $-(P^2+2PQ\sin\psi)$ in (49) is much smaller than the positive term $(1-S-P^2)\sin^2\psi$. To verify this statement, note that in a nearly polar orbit, $S\approx 1$, $Q\approx 1$, $P<<1$, and near a pole $|\sin\psi|\approx 1$. Then from (32.3) of V1966

$$P=\left(1-\frac{c^2}{a_0P_0}Su\right)^{-1}\frac{\delta}{p_o}u(1-S)\approx\frac{7}{6400}(1-S), \tag{49.1}$$

so that

$$|P^2+2PQ\sin\psi|\approx\frac{7}{3200}(1-S) \tag{49.2}$$

and

$$(1-S-P^2)\sin^2\psi\approx 1-S \tag{49.3}$$

Thus Eq. (49) gives no trouble near a pole.

In rectangular coordinates we find from (1), (24), (44), and (47)

$$X = (\rho^2+c^2)^{\frac{1}{2}}\left[H_1\cos\Omega'\cos\psi-H_1^{-1}\sqrt{1-S}\,\mathrm{sgn}\alpha_3\sin\Omega'(H_2+H_3\sin\psi)\right] \tag{50.1}$$

$$Y = (\rho^2+c^2)^{\frac{1}{2}}\left[H_1\sin\Omega'\cos\psi+H_1^{-1}\sqrt{1-S}\,\mathrm{sgn}\alpha_3\cos\Omega'(H_2+H_3\sin\psi)\right] \tag{50.2}$$

and

$$Z = \rho\eta-\delta\ , \tag{50.3}$$

from (1.2) of V1966. Here

$$H_1 = \pm 2^{-\frac{1}{2}}\left[1+S+(1-S)\sqrt{1-r^2\delta^2}\,\right]^{\frac{1}{2}} \tag{51.1}$$

$$H_2 = \tfrac{1}{2}Q\left(\sqrt{1-r\delta}-\sqrt{1-r\delta}\,\right) \tag{51.2}$$

$$H_3 = \tfrac{1}{2}\left[(1+P)\sqrt{1-r\delta}+(1-P)\sqrt{1+r\delta}\,\right] \tag{51.3}$$

and

$$\Omega'=\beta_3-c^2\alpha_3(-2\alpha_1)^{-\frac{1}{2}}(A_3 v+\sum_{k=1}^{4}A_{3k}\sin kv)$$

$$+\alpha_3\alpha_2^{-1}u^{\frac{1}{2}}(B_3\psi-\frac{3}{4}C_1C_2Q\cos\psi+\frac{3}{32}C_2{}^2Q^2\sin 2\psi), \qquad (51.4)$$

from Eq.(150) of V1966. Like Eqs.(10) these equations contain no singularities, even for a polar orbit. Moreover they hold for all oribts.

For an exactly polar orbit we have S=1, P=0, Q=1, $\alpha_3=0$, and $\Omega'=\beta_3$. The X and Y equations then become

$$X+iY=(\rho^2+c^2)^{\frac{1}{2}}\cos\psi\exp i\beta_3, \qquad (52)$$

as for the case $J_3=0$ of Eq. (11). The Z equation, however, is $Z=\rho\eta-\delta$, where $\delta=(r_e/2)J_2^{-1}|J_3|$, so that the orbit is still changed by the J_3.

7. Velocity Components, with J_3 Accounted for

From Eqs. (50.1) and (50.2)

$$X+iY=(\rho^2+c^2)^{\frac{1}{2}}[H_1\cos\psi+iH_1^{-1}(1-S)^{\frac{1}{2}}\operatorname{sgn}\alpha_3(H_2+H_3\sin\psi)]\exp i\Omega' \qquad (53)$$

Logarithmic differentiation of (53), with multiplication of the result by X+iY, gives

$$\dot{X}+i\dot{Y}=(\frac{\rho\dot{\rho}}{\rho^2+c^2}+i\dot{\Omega}')(X+iY)+(\rho^2+c^2)^{\frac{1}{2}}[-H_1\sin\psi+iH_1^{-1}(1-S)^{\frac{1}{2}}\operatorname{sgn}\alpha_3H_3\cos\psi]\dot{\psi}\exp i\Omega', \qquad (54)$$

so that

$$\dot{X}=\frac{\rho\dot{\rho}}{\rho^2+c^2}X-Y\dot{\Omega}'+(\rho^2+c^2)^{\frac{1}{2}}[-H_1\sin\psi\cos\Omega'-H_1^{-1}(1-S)^{\frac{1}{2}}\operatorname{sgn}\alpha_3H_3\cos\psi\sin\Omega']\dot{\psi} \qquad (54.1)$$

$$\dot{Y}=\frac{\rho\dot{\rho}}{\rho^2+c^2}\,Y+X\dot{\Omega}'+(\rho^2+c^2)^{\frac{1}{2}}[-H_1\sin\psi\sin\Omega'+H_1^{-1}(1-S)^{\frac{1}{2}}\operatorname{sgn}\alpha_3 H_3\cos\psi\cos\Omega']\dot{\psi} \tag{54.2}$$

Also

$$\dot{Z} = \eta\dot{\rho} + \rho\dot{\eta} \tag{54.3}$$

by (15). Eqs. (16) and (23) still hold, so that the equations for $\dot{\rho}$ and $\dot{v}$ are as for $J_3=0$.

For $\dot{\Omega}'$ we find from (51.4)

$$\dot{\Omega}'=-c^2\alpha_3(-2\alpha_1)^{-\frac{1}{2}}(A_3+\sum_{k=1}^{4}kA_{3k}\cos kv)\dot{v}$$

$$+\alpha_3\alpha_2^{-1}u^{\frac{1}{2}}(B_3+\tfrac{3}{4}C_1C_2Q\sin\psi+\tfrac{3}{16}C_2^2Q^2\cos 2\psi)\dot{\psi} \tag{55}$$

The new expression for $\dot{\psi}$ is still lacking. From p. 14 of Bonavito 1966, we find

$$\dot{\eta} = Q\cos\psi\quad \dot{\psi} = \frac{Q}{\rho^2+c^2\eta^2}\left(\frac{\mu p_0}{u}\right)^{\frac{1}{2}}(1-C_1\eta-C_2\eta^2)^{\frac{1}{2}}\cos\psi\,, \tag{56}$$

so that

$$\dot{\psi} = \left(\frac{\mu p_0}{u}\right)^{\frac{1}{2}}\frac{(1-C_1\eta-C_2\eta^2)^{\frac{1}{2}}}{\rho^2+c^2\eta^2} \tag{57}$$

Here

$$\frac{1}{u} = 1+\frac{c^2}{a_0p_0}(1-S)+\frac{\left(\frac{2\delta}{p_0}\right)^2(1-S)\left(1-\frac{c^2}{a_0p_0}S\right)}{\left[1+\frac{c^2}{a_0p_0}(1-2S)\right]^2}+O(J_2^4), \tag{58}$$

$$c^2 = c_0^2-\delta^2 = r_e^2J_2-\tfrac{1}{4}r_e^2J_3^2J_2^{-2} \tag{59}$$

The equations of this section reduce to those of Section 5, if J_3 is equated to zero.

8. The Improved Algorithm for the Spheroidal Model, with J_3

Begin with Section 12 of V1966 and follow it through the third line on p.45, viz, $\eta = P+Q \sin \Psi$. Instead of then calculating E_2' and E_3', however, replace that calculation with Eqs. (42), (50), and (51) of the present paper. This changed procedure not only simplifies the calculation of X, Y, and Z for near-polar orbits but bypasses the right ascension in all cases. To calculate the velocities $\dot{X}$, $\dot{Y}$, and $\dot{Z}$, use Eqs. (54) through (59) of the present paper.

9. References

Bonavito, N. L., 1962, NASA Technical Note D-1177

Bonavito, N. L., 1966, NASA Technical Note D-3562

Izsak, I., 1960, Smithsonian Institution Astrophysical Observatory Research in Space Science, Special Report No. 52.

Izsak, I., 1963, Smithsonian Contributions to Astrophysics, Vol. 6, Research in Space Science, 81-107

Vinti, J. P., 1959, J. Research Nat. Bureau Standards, 63B, 105-116

Vinti, J. P., 1961a, J. Research Nat. Bureau Standards, 65B, 169-201

Vinti, J. P., 1961b, Astron. J., 66, 514-516

Vinti, J. P., 1966, J. Research Nat. Bureau Standards, 70B, 17-46

Walden, H. and Watson, S., 1967, NASA Technical Note TN-D-4088

Walden, H., 1967, NASA Technical Note TN D-3803

Appendix I

Algorithm for Satellite Position Vector and Velocity, if the Potential $V = -\mu\rho(\rho^2+c^2\eta^2)^{-\frac{1}{2}}(J_3 = 0)$

Given

$$\mu,\ r_e,\ J_2\ a,\ e,\ I,\ \beta_1,\ \beta_2,\ \beta_3$$

Compute once for each orbit:

$$c_0^2=r_e^2J_2,\quad \eta_0=\sin I,\quad p=a(1-e^2),\quad D=(ap-c_0^2)(ap-c_0^2\eta_0^2)+4a^2c_0^2\eta_0^2$$

$$D'=D+4a^2c_0^2(1-\eta_0^2),\quad A=-2ac_0^2D^{-\frac{1}{2}}(1-\eta_0^2)(ap-c_0^2\eta_0^2)<0,\quad B=c_0^2\eta_0^2D^{-1}D'$$

$$b_1=-\tfrac{1}{2}A>0,\quad b_2=B^{\frac{1}{2}},\quad a_0=a+b_1>a,\quad p_0=-c_0^2a_0^{-1}(1-\eta_0^{-1})+aa_0^{-1}pD^{-\frac{1}{2}}D'$$

$$\alpha_2=(\mu p_0)^{\frac{1}{2}},\quad \alpha_3=\alpha_2\left(1-\frac{c_0^2\eta_0^2}{a_0p_0}\right)\cos I,\quad \eta_2^{-2}=\frac{c_0^2D}{apD'},\quad q=\eta_0\eta_2^{-1}$$

$$\alpha_2'=\alpha_2\left(1+\frac{c_0^2}{a_0p_0}\cos^2 I\right)^{\frac{1}{2}}$$

Also, with $R_n(x) \equiv x^nP_n(x^{-1})$,

compute

$$A_1=(1-e^2)^{\frac{1}{2}}p\sum_{n=2}^{\infty}\left(\frac{b_2}{p}\right)^nP_n\left(\frac{b_1}{b_2}\right)R_{n-2}[(1-e^2)^{\frac{1}{2}}]$$

$$A_2=(1-e^2)^{\frac{1}{2}}p^{-1}\sum_{n=0}^{\infty}(b_2/p)^nP_n(b_1/b_2)R_n[(1-e^2)^{\frac{1}{2}}],\ \text{where}$$

$$D_{2i}=\sum_{n=0}^{i}(-1)^{i-n}(c_0/p)^{2i-2n}(b_2/p)^{2n}P_{2n}(b_1/b_2)$$

$$D_{2i+1}=\sum_{n=0}^{i}(-1)^{i-n}(c_0/p)^{2i-2n}(b_2/p)^{2n+1}P_{2n+1}(b_1/b_2)$$

$$A_3=(1-e^2)^{\frac{1}{2}}p^{-3}\sum_{m=0}^{\infty}D_mR_{m+2}[(1-e^2)^{\frac{1}{2}}]$$

$$B_1 = 2\pi^{-1}q^{-2}[K(q)-E(q)]=\frac{1}{2}+\frac{3}{16}q^2+\frac{15}{128}q^4+\frac{175}{2048}q^6+.....$$

$$B_2 = 2\pi^{-1}K(q) = 1+\frac{1}{4}q^2+\frac{9}{64}q^4+\frac{25}{256}q^6+....$$

$$a_0' = a_0+A_1+c_0^2\eta_0^2A_2B_1B_2^{-1}$$

$$\gamma_m = \frac{(2m)!}{2^{2m}(m!)^2}\sum_{n=1}^{m-1}\frac{(2n)!\eta_0^{2n}}{2^{2n}(n!)^2}$$

$$B_3 =1-(1-\eta_2^{-2})^{\frac{1}{2}}-\sum_{m=2}^{\infty}\gamma_m\eta_2^{-2m}$$

$$A_{11} = \frac{3}{4}(1-e^2)^{\frac{1}{2}}p^{-3}e(-2b_1b_2^2p+b_2^4) \quad A_{12}=\frac{3}{32}p^{-3}(1-e^2)^{\frac{1}{2}}b_2^4e^2$$

$$A_{21}=(1-e^2)^{\frac{1}{2}}p^{-1}e[b_1p^{-1}+(3b_1^2-b_2^2)p^{-2}-\frac{9}{2}b_1b_2^2(1+\frac{e^2}{4})p^{-3}+\frac{3}{8}b_2^4(4+3e^2)p^{-4}]$$

$$A_{22}=(1-e^2)^{\frac{1}{2}}p^{-1}[\frac{e^2}{8}(3b_1^2-b_2^2)p^{-2}-\frac{9}{8}e^2b_1b_2^2p^{-3}+\frac{3}{32}b_2^4(6e^2+e^4)p^{-4}]$$

$$A_{23}=(1-e^2)^{\frac{1}{2}}p^{-1}\frac{e^3}{8}(-b_1b_2^2p^{-3}+b_2^4p^{-4}), \quad A_{24}=\frac{3}{256}(1-e^2)^{\frac{1}{2}}p^{-5}b_2^4e^4$$

$$A_{31}=(1-e^2)^{\frac{1}{2}}p^{-3}e[2+b_1p^{-1}(3+\frac{3}{4}e^2)-p^{-2}(\frac{1}{2}b_2^2+c_0^2)(4+3e^2)]$$

$$A_{32}=(1-e^2)^{\frac{1}{2}}p^{-3}\left[\frac{e^2}{4}+\frac{3}{4}b_1p^{-1}e^2-p^{-2}\left(\frac{e^4}{4}+\frac{3}{2}e^2\right)\left(\tfrac{1}{2}b_2{}^2+c_0{}^2\right)\right]$$

$$A_{33}=(1-e^2)^{\frac{1}{2}}p^{-3}e^3\left[\tfrac{1}{12}b_1p^{-1}-\tfrac{1}{3}p^{-2}\left(\tfrac{1}{2}b_2{}^2+c_0{}^2\right)\right],\quad A_{34}=-\tfrac{1}{32}(1-e^2)^{\frac{1}{2}}p^{-5}e^4\left(\tfrac{1}{2}b_2{}^2+c_0{}^2\right)$$

$$2\pi\nu_1=a_0'^{-1}\left(\frac{\mu}{a_0}\right)^{\frac{1}{2}},\qquad 2\pi\nu_2=a_0'^{-1}\alpha_2'A_2B_2^{-1},\qquad e'=a_0^{-1}ae<e$$

$$\lambda_1=\beta_1-c_0{}^2\beta_2\alpha_2^{-1}\eta_0{}^2B_1B_2^{-1},\quad \lambda_2=\beta_1+\beta_2\alpha_2^{-1}(a_0+A_1)A_2^{-1}$$

$$\lambda_3=\left(\frac{\mu}{a_0}\right)^{-\frac{1}{2}}\alpha_2'A_2B_2^{-1},\qquad \lambda_4=a_0^{-1}(A_1+c_0{}^2\eta_0{}^2A_2B_1B_2^{-1})$$

$$\lambda_5=c_0{}^2\left(\frac{\mu}{a_0}\right)^{\frac{1}{2}}\alpha_2'\eta_0{}^4,\qquad \lambda_6=\left(\frac{\mu}{a_0}\right)^{-\frac{1}{2}}\alpha_2'B_2^{-1},\qquad \lambda_7=\tfrac{1}{8}q^2B_2^{-1}$$

For <u>each point at time t</u>, now compute

1) $M_s=2\pi\nu_1(t+\lambda_1)\qquad \psi_s=2\pi\nu_2(t+\lambda_2)$

2) Solve for E_0: $M_s+E_0-e'\sin(M_s+E_0)=M_s$

3) To find v_0: Place $E=M_s+E_0$ in the anomaly connections

$$\cos v=(1-e\cos E)^{-1}(\cos E-e),\quad \sin v=(1-e\cos E)^{-1}(1-e^2)^{\frac{1}{2}}\sin E$$

and solve for $v=M_s+v_0$

4) $\psi_0=\lambda_3v_0$

5) Compute $M_1=-\lambda_4v_0+\frac{1}{4}\lambda_5\sin(2\psi_s+2\psi_0)$

6) Then $E_1=[1-e'\cos(M_s+E_0)]^{-1}M_1-\frac{1}{2}e'[1-e'\cos(M_s+E_0)]^{-3}M_1^2\sin(M_s+E_0)$

7) Place $E=M_s+E_0+E_1$ in the anomaly connections and solve for $v=M_s+v_0+v_1$

8) Then $\psi_1=\lambda_6[A_2v_1+\sum_{k=1}^{2}A_{2k}\sin(kM_s+kv_0)]+\lambda_7\sin(2\psi_s+2\psi_0)$

9) Compute

$$M_2=-a_0^{-1}[A_1v_1+\sum_{k=1}^{2}A_{1k}\sin(kM_s+kv_0)+\lambda_5\{B_1\psi_1-\frac{1}{2}\psi_1\cos(2\psi_s+2\psi_0)-\frac{q^2}{8}\sin(2\psi_s+2\psi_0)+\frac{q^4}{64}\sin(4\psi_s+4\psi_0)\}]$$

10) Then $E_2=[1-e'\cos(M_s+E_0+E_1)]^{-1}M_2$

11) Place $E=M_s+E_0+E_1+E_2$ in the anomaly connections to find $v=M_s+v_0+v_1+v_2$.

12) Then $\psi_2=\lambda_6[A_2v_2+A_{21}v_1\cos(M_s+v_0)+2A_{22}v_1\cos(2M_s+2v_0)+A_{23}\sin(3M_s+3v_0)+A_{24}\sin(4M_s+4v_0)]+2\lambda_7[\psi_1\cos(2\psi_s+2\psi_0)+\frac{3q^2}{8}\sin(2\psi_s+2\psi_0)-\frac{3q^2}{64}\sin(4\psi_s+4\psi_0)]$

Then

$$E=M_s+E_0+E_1+E_2,\quad v=M_s+v_o+v_1+v_2,\quad \psi=\psi_s+\psi_0+\psi_1+\psi_2$$

13) $\rho=a(1-e\cos E)=(1+e\cos v)^{-1}p,\quad r=r_0\sin\psi$

14) $\Omega'=\beta_3+\alpha_3\alpha_2'^{-1}(B_3\psi+\frac{3}{32}\eta_0^2\eta_2^{-4}\sin2\psi)-c^2\alpha_3\left(\frac{\mu}{a_0}\right)^{-\frac{1}{2}}(A_3v+\sum_{k=1}^{4}A_{3k}\sin kv)$

Then the rectangular coordinates are given by

15) $X=(\rho^2+c_0^2)^{\frac{1}{2}}(\cos\Omega'\cos\psi-\sin\Omega'\cos I\sin\psi)$

16) $Y=(\rho^2+c_0^2)^{\frac{1}{2}}(\sin\Omega'\cos\psi+\cos\Omega'\cos I\sin\psi)$

17) $Z=\rho\eta$

To find the velocity components, compute

18) $\dot{v}=\frac{a}{\rho}\left[\frac{\mu(1-e^2)}{a_0}\right]^{\frac{1}{2}}\frac{(\rho^2+A\rho+B)^{\frac{1}{2}}}{\rho^2+c_0^2\eta^2}$

19) $\dot{\rho}=\frac{e}{p}\rho^2\sin v\,\dot{v}$

20) $\dot{\psi}=\frac{\alpha_2'(1-q^2\sin^2\psi)^{\frac{1}{2}}}{\rho^2+c_0^2\eta^2}$

21) $\dot{\Omega}'=\alpha_3\alpha_2'^{-1}(B_3+\frac{3}{16}\eta_0^2\eta_2^{-4}\cos 2\psi)\dot{\psi}$

$$-c^2\alpha_3\left(\frac{\mu}{a_0}\right)^{-\frac{1}{2}}(A_3+\sum_{k=1}^{4}kA_{3k}\cos kv)\dot{v}$$

Then

22) $\dot{X}=\frac{\rho\dot{\rho}}{\rho^2+c_0^2}X-Y\dot{\Omega}'+(\rho^2+c_0^2)^{\frac{1}{2}}(-\sin\psi\cos\Omega'-\cos I\cos\psi\sin\Omega')\dot{\psi}$

23) $\dot{Y}=\frac{\rho\dot{\rho}}{\rho^2+c_0^2}Y+X\dot{\Omega}'+(\rho^2+c_0^2)^{\frac{1}{2}}(-\sin\psi\sin\Omega'+\cos I\cos\psi\cos\Omega')\dot{\psi}$

24) $\dot{Z}=\eta\dot{\rho}+\eta_0\rho\cos\psi\,\dot{\psi}$

Appendix II

Algorithm for Satellite Position Vector and Velocity, if the Potential $V = -\mu(\rho+\eta\delta)(\rho^2+c^2\eta^2)^{-1}(J_3\neq 0)$

Given $\mu, r_e, J_2, J_3, a, e, S, \beta_1, \beta_2, \beta_3$

Compute once for each orbit

$$c_0^2 = r_e^2 J_2, \quad \delta = \frac{1}{2} r_e \frac{|J_3|}{J_2}, \quad c^2 = c_0^2 - \delta^2, \quad p = a(1-e^2)$$

$$A = \frac{-2ac^2(ap-c^2S)(1-S) + \frac{8a^2c^2}{p}\delta^2\left\{1+\frac{c^2}{ap}(3S-2)\right\}S(1-S)}{(ap-c^2)(ap-c^2S)+4a^2c^2S+\frac{4c^2}{p^2}\delta^2(3ap-4a^2-c^2)S(1-S)}$$

$$B = c^2+(2a)^{-1}(ap-c^2)A, \quad b_1 = -\frac{1}{2}A, \quad a_0 = a+b_1, \quad b_2 = B^{\frac{1}{2}}$$

$$p_0 = a_0^{-1}(B+ap-2Aa-c^2), \quad \alpha_2 = (\mu p_0)^{\frac{1}{2}}, \quad u \text{ from}$$

$$u^{-1} = 1+\frac{c^2}{a_0p_0}(1-S)+\frac{\left(\frac{2\delta}{p_0}\right)^2(1-S)\left(1-\frac{c^2}{a_0p_0}S\right)}{\left[1+\frac{c^2}{a_0p_0}(1-2S)\right]^2}, \quad C_2 = \frac{c^2}{a_0p_0}u, \quad \alpha_3 = \pm\alpha_2(1-Su^{-1})^{\frac{1}{2}}$$

\+ for direct orbit
\- for retrograde

$$C_1 = \left(1-\frac{c^2}{a_0p_0}Su\right)^{-1}\frac{2\delta}{p_0}u\left(1-\frac{c^2}{a_0p_0}u\right), \quad P = \left(1-\frac{c^2}{a_0p_0}Su\right)^{-1}\frac{\delta}{p_0}u(1-S),$$

With $R_n(x) = x^n P_n(x^{-1})$, compute

$$A_1 = (1-e^2)^{\frac{1}{2}}p\sum_{n=2}^{\infty}(b_2/p)^n P_n(b_1/b_2)R_{n-2}[(1-e^2)^{\frac{1}{2}}]$$

$$A_2 = (1-e^2)^{\frac{1}{2}}p^{-1}\sum_{n=0}^{\infty}(b_2/p)^n P_n(b_1/b_2)R_n[(1-e^2)^{\frac{1}{2}}]$$

$A_3=(1-e^2)^{\frac{1}{2}}p^{-3}\sum_{m=0}^{\infty} D_m R_{m+2}[(1-e^2)^{\frac{1}{2}}]$, where

$$D_{2i} = \sum_{n=0}^{i} (-1)^{i-n}(c/p)^{2i-2n}(b_2/p)^{2n}P_{2n}(b_1/b_2)$$

$$D_{2i+1}= \sum_{n=0}^{i} (-1)^{i-n}(c/p)^{2i-2n}(b_2/p)^{2n+1}P_{2n+1}(b_1/b_2)$$

$A_{11}= \frac{3}{4}(1-e^2)^{\frac{1}{2}}p^{-3}e(-2b_1b_2{}^2p+b_2{}^4)$ $\quad A_{12}= \frac{3}{32}p^{-3}(1-e^2)^{\frac{1}{2}}b_2{}^4e^2$

$A_{21}= (1-e^2)^{\frac{1}{2}}p^{-1}e[b_1p^{-1}+(3b_1{}^2-b_2{}^2)p^{-2}- \frac{9}{2}b_1b_2{}^2(1+\frac{e^2}{4})p^{-3}+ \frac{3}{8}b_2{}^4(4+3e^2)p^{-4}]$

$A_{22}=(1-e^2)^{\frac{1}{2}}p^{-1}[\frac{e^2}{8}(3b_1{}^2-b_2{}^2)p^{-2}- \frac{9}{8}e^2b_1b_2{}^2p^{-3}+ \frac{3}{32}b_2{}^4(6e^2+e^4)p^{-4}]$

$A_{23}=(1-e^2)^{\frac{1}{2}}p^{-1}\frac{e^3}{8}(-b_1b_2{}^2p^{-3}+b_2{}^4p^{-4})$, $\quad A_{24}= \frac{3}{256}(1-e^2)^{\frac{1}{2}}p^{-5}b_2{}^4e^4$

$A_{31}=(1-e^2)^{\frac{1}{2}}p^{-3}e[2+b_1p^{-1}(3+\frac{3}{4}e^2)-p^{-2}(\frac{1}{2}b_2{}^2+c^2)(4+3e^2)]$

$A_{32}=(1-e^2)^{\frac{1}{2}}p^{-3}[\frac{e^2}{4}+\frac{3}{4}b_1p^{-1}e^2-p^{-2}(\frac{e^4}{4}+\frac{3}{2}e^2)(\frac{1}{2}b_2{}^2+c^2)]$

$A_{33}=(1-e^2)^{\frac{1}{2}}p^{-3}c^3[\frac{1}{12}b_1p^{-1}-\frac{1}{3}p^{-2}(\frac{1}{2}b_2{}^2+c^2)]$, $\quad A_{34}= -\frac{1}{32}(1-e^2)^{\frac{1}{2}}p^{-5}(\frac{1}{2}b_2{}^2+c^2)e^4$

$Q = (P^2+S)^{1/2}$

$B_2=1-\frac{1}{2}C_1P+(\frac{3}{8}C_1{}^2+\frac{1}{2}C_2)(\frac{1}{2}Q^2)+\frac{9}{64}C_2{}^2Q^4+0(J_2{}^3)$

$B_1 = \frac{1}{2}Q^2+P^2-\frac{3}{4}C_1PQ^2+\frac{3}{64}(4C_2+3C_1{}^2)Q^4+\frac{15}{128}C_2{}^2Q^6+0(J_2{}^3)$

$B_3 = -\frac{1}{2}C_2 - \frac{3}{8}C_1^2 - \frac{3}{8}C_2^2(1+\frac{1}{2}Q^2)+O(J_2^3)$,

$B_{11} = -2PQ+\frac{3}{8}C_1Q^3$, $B_{12} = -(\frac{Q^2}{4}+\frac{1}{8}C_2Q^4)$, $B_{13} = -C_1\frac{Q^3}{24}$, $B_{14} = C_2\frac{Q^4}{64}$

$B_{21} = -C_2PQ+\frac{9}{16}C_1C_2Q^3+\frac{1}{2}C_1Q$, $B_{22} = -\frac{1}{32}[(4C_2+3C_1^2)Q^2+3C_2^2Q^4]$

$B_{23} = -\frac{1}{16}C_1C_2Q^3$, $B_{24} = \frac{3}{256}C_2^2Q^4$, $r = 2(1-C_2S)^{-1}up_0^{-1}$

$a_0' = a_0+A_1+c^2A_2B_1'B_2^{-1}$, $2\pi\nu_1 = \left(\frac{\mu}{a_0}\right)^{\frac{1}{2}}(a_0')^{-1}$, $2\pi\nu_2 = \alpha_2 u^{-\frac{1}{2}}A_2B_2^{-1}(a_0')^{-1}$,

$e' = a\ e\ a_0^{-1}$

$\lambda_1 = \beta_1 - c^2\beta_2\alpha_2^{-1}B_1'B_2^{-1}$, $\lambda_2 = \beta_1+\beta_2\alpha_2^{-1}(a_0+A_1)A_2^{-1}$, $\lambda_3 = \left(\frac{\mu}{a_0}\right)^{-\frac{1}{2}}\alpha_2 u^{-\frac{1}{2}}A_2B_2^{-1}$

$\lambda_4 = a_0^{-1}(A_1+c^2A_2B_1'B_2^{-1})$, $\lambda_5 = \left(\frac{\mu}{a_0}\right)^{\frac{1}{2}}c^2\alpha_2^{-1}u^{\frac{1}{2}}$, $\lambda_6 = \left(\frac{\mu}{a_0}\right)^{-\frac{1}{2}}\alpha_2 u^{-\frac{1}{2}}B_2^{-1}$,

$H_1 = 2^{-\frac{1}{2}}[1+S+(1-S)(1-r^2\delta^2)^{\frac{1}{2}}]^{\frac{1}{2}}$ $H_2 = \frac{1}{2}Q[(1-r\delta)^{\frac{1}{2}}-(1+r\delta)^{\frac{1}{2}}]$,

$H_3 = \frac{1}{2}[(1+P)(1-r\delta)^{\frac{1}{2}}+(1-P)(1+r\delta)^{\frac{1}{2}}$

Compute for each point

1) $M_s = 2\pi\nu_1(t+\lambda_1)$, $\psi_s = 2\pi\nu_2(t+\lambda_2)$

2) Solve for E_0: $M_s+E_0-e'\sin(M_s+E_0) = M_s$

3) To find v_0: Place $E = M_s+E_0$ in the anomaly connections

4) $\psi_0 = \lambda_3 v_0$

5) Compute $M_1 = -\lambda_4 v_0 - a_0^{-1}\lambda_5 B_{12}\sin(2\psi_s+2\psi_0)$

6) Then $E_1=[1-e'\cos(M_s+E_0)]^{-1} M_1 - \frac{1}{2}e'[1-e'\cos(M_s+E_0)]^{-3} M_1^2 \sin(M_s+E_0)$

7) Place $E=M_s+E_0+E_1$ in the anomaly connections and solve for $v=M_s+v_0+v_1$

8) Then $\psi_1=\lambda_6[A_2 v_1+\sum_{k=1}^{2} A_{2k}\sin(kM_s+kv_0)]-B_{21}B_2^{-1}\cos(\psi_s+\psi_0)-B_{22}B_2^{-1}\cdot$

$\sin(2\psi_s+2\psi_0)$

9) Compute $M_2=-a_0^{-1}[A_1 v_1+\sum_{k=1}^{2} A_{1k}\sin(kM_s+kv_0)+\lambda_5\{B_1'\psi_1+B_{11}\cos(\psi_s+\psi_0)$

$+2B_{12}\psi_1\cos(2\psi_s+2\psi_0)+B_{13}\cos(3\psi_s+3\psi_0)+B_{14}\sin(4\psi_s+4\psi_0)\}]$

10) Then $E_2=[1-e'\cos(M_s+E_0+E_1)]^{-1}M_2$

11) Place $E=M_s+E_0+E_1+E_2$ in the anomaly connections to find

$v=M_s+v_0+v_1+v_2$

12) Then $\psi_2=\lambda_6[A_2 v_2+A_{21}v_1\cos(M_s+v_0)+2A_{22}v_1\cos(2M_s+2v_0)$

$+A_{23}\sin(3M_s+3v_0)+A_{24}\sin(4M_s+4v_0)]-B_2^{-1}[-B_{21}\psi_1\sin(\psi_s+\psi_0)$

$+2B_{22}\psi_1\cos(2\psi_s+2\psi_0)+B_{23}\cos(3\psi_s+3\psi_0)+B_{24}\sin(4\psi_s+4\psi_0)]$

Then $E=M_s+E_0+E_1+E_2$, $v=M_s+v_0+v_1+v_2$, $\psi=\psi_s+\psi_0+\psi_1+\psi_2$

13) $\rho=a(1-\cos E)=(1+e\cos v)^{-1}p$, $\eta=P+Q\sin\psi$

14) $\Omega'=\beta_3-c^2\alpha_3\left(\frac{\mu}{a_0}\right)^{-\frac{1}{2}}(A_3 v+\sum_{k=1}^{4} A_{3k}\sin kv)+\alpha_3\alpha_2^{-1}u^{\frac{1}{2}}(B_3\psi-\frac{3}{4}C_1C_2Q\cos\psi$

$+\frac{3}{32}C_2^2Q^2\sin 2\psi)$

Then if sgn $\alpha_3 = \pm 1$ for direct or retrograde orbits respectively, the rectangular coordinates are

15) $$X=(\rho^2+c^2)^{\frac{1}{2}}[H_1\cos\Omega'\cos\psi-H_1^{-1}(1-S)^{\frac{1}{2}}\mathrm{sgn}\alpha_3(H_2+H_3\sin\psi)\sin\Omega']$$

16) $$Y=(\rho^2+c^2)^{\frac{1}{2}}[H_1\sin\Omega'\cos\psi+H_1^{-1}(1-S)^{\frac{1}{2}}\mathrm{sgn}\alpha_3(H_2+H_3\sin\psi)\cos\Omega']$$

17) $$Z=\rho\eta-\delta$$

Velocity Components

18) $$\dot{v}=\frac{a}{\rho}\frac{\mu(1-e^2)}{a_0}^{\frac{1}{2}}\frac{(\rho^2+A\rho+B)^{\frac{1}{2}}}{\rho^2+c^2\eta^2}$$

19) $$\dot{\rho}=\frac{e}{p}\rho^2\sin v\,\dot{v}$$

20) $$\dot{\psi}=\left(\frac{\mu p_0}{u}\right)^{\frac{1}{2}}(\rho^2+c^2\eta^2)^{-1}(1+C_1\eta-C_2\eta^2)^{\frac{1}{2}}$$

21) $$\dot{\Omega}'=-c^2\alpha_3\left(\frac{\mu}{a_0}\right)^{\frac{-1}{2}}(A_3+\sum_{k=1}^{4}kA_{3k}\cos kv)\dot{v}$$

$$+\alpha_3\alpha_2^{-1}u^{\frac{1}{2}}(B_3+\frac{3}{4}C_1C_2Q\sin\psi+\frac{3}{16}C_2^2Q^2\cos 2\psi)\dot{\psi}$$

22) $$\dot{X}= \frac{\rho\dot{\rho}}{\rho^2+c^2} X-Y\dot{\Omega}'+(\rho^2+c^2)^{\frac{1}{2}}[-H_1 \sin\psi\cos\Omega'-H_1^{-1}(1-S)^{\frac{1}{2}}\operatorname{sgn}\alpha_3 H_3\cos\psi\sin\Omega']\dot{\psi}$$

23) $$\dot{Y}= \frac{\rho\dot{\rho}}{\rho^2+c^2} Y+X\dot{\Omega}'+(\rho^2+c^2)^{\frac{1}{2}}[-H_1 \sin\psi\sin\Omega'+H_1^{-1}(1-S)^{\frac{1}{2}}\operatorname{sgn}\alpha_3 H_3\cos\psi\cos\Omega']\dot{\psi}$$

24) $$\dot{Z}= \eta\dot{\rho} + \rho Q\cos\psi\ \dot{\psi}$$

EXPANSION FORMULAE IN CANONICAL TRANSFORMATIONS DEPENDING ON A SMALL PARAMETER

By Ahmed Aly Kamel
Stanford University
Stanford, California

EXPANSION FORMULAE IN CANONICAL TRANSFORMATIONS DEPENDING ON A SMALL PARAMETER

by

Ahmed Aly Kamel

Stanford University, Stanford, California

ABSTRACT

The theory of perturbation based on Lie transforms is considered. Deprit's equation is reduced to a form which enables us to generate simplified general recursion formulae. These expansions are then modified to speed up the implementation of such perturbation theory in the computerized symoblic manipulation.

1. INTRODUCTION

If a system is described by a Hamiltonian depending on a small parameter, then canonical transformation can sometimes be obtained using a von Zeipel generating function (See for instance Brouwer and Clemence 1961). In such a case, the transformation is implicit because the generating function is in mixed variables (the old coordinates and the new momenta).

The shortcomings of von Zeipel's method, when the generating function itself depends on a small parameter, were felt by Breakwell and Pringle (1966), and Deprit (1966), when they used a von Zeipel generating function to remove the short period terms from the Hamiltonian of

a particle in the neighborhood of the triangular points in the restricted problem of three bodies. Breakwell (See Schechter 1968) recognized, after comparing with Deprit et al (1967), that the long period part of the second order Hamiltonian derived in mixed variables was misleading, and that it was possible to obtain a different representation in terms of new variables only. Using this suggestion, Schechter (1968) obtained a more valid second order expression. Deprit (1968) attacked the problem using Lie transforms and extended the expansion to include higher orders. In this paper Deprit's recursive algorithm is reduced to a form which enables us to generate simplified and modified general formulae (Section 3 and Section 4).

2. BACKGROUND

A Lie transform may be defined by the differential equations

$$\frac{dx}{d\epsilon} = W_X(x,X,t;\epsilon) \tag{2.1a}$$

$$\frac{dX}{d\epsilon} = -W_x(x,X,t;\epsilon) \tag{2.1b}$$

$$\frac{dt}{d\epsilon} = 0 \tag{2.1c}$$

$$\frac{dF}{d\epsilon} = W_t(x,X,t;\epsilon) \tag{2.1d}$$

with the initial conditions $x = y$, $X = Y$, $t = t$, and $F = 0$ at $\epsilon = 0$. The foregoing equations define a canonical transformation. This can be shown as follows: for any ϵ, the differentials dx, dX, δx, δX, and

δF produced by the initial changes dy, dY, dt, δy, and δY satisfy the equation*

$$\frac{d}{d\epsilon}\left[dx.\delta X - dX.\delta x - dt\delta F\right] = 0 \quad . \tag{2.2}$$

Hence, $dx.\delta X - dX.\delta x - dt\delta F$ is independent of ϵ and equals its value at $\epsilon = 0$, so that for $F = H(x,X,t;\epsilon) - K(y,Y,t;\epsilon)$

$$\dot{x}.\delta X - \dot{X}.\delta x - \delta H = \dot{y}.\delta Y - \dot{Y}.\delta y - \delta K. \tag{2.3}$$

Therefore, if x and X satisfy the canonical equations

$$\dot{x} = H_X, \ \dot{X} = -H_x \ , \tag{2.4}$$

then, also y and Y satisfy the canonical equations

$$\dot{y} = K_Y \ , \ \dot{Y} = -K_y \quad . \tag{2.5}$$

Now, take any indefinitely differentiable function $f(x,X,t;\epsilon)$ that can be expressed in terms of x,X,t and ϵ as a power series in ϵ, in the form

$$f(x,X,t;\epsilon) = \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!}\left[\frac{\partial^n}{\partial\epsilon^n} f(x,X,t;\epsilon)\right]_{\epsilon=0} = \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!} f_n(x,X,t) \tag{2.6}$$

---

*Notice that

$$\frac{d}{d\epsilon}dx = dW_X = W_{Xx}.dx + W_{XX}.dX + W_{Xt}dt \ ,$$

$$\frac{d}{d\epsilon}\delta x = \delta W_X = W_{Xx}.\delta x + W_{XX}.\delta X \ , \text{ etc.}$$

then, in terms of $y,Y,t,$ and ϵ as a power series in ϵ, it takes the form

$$f(x,X,t;\epsilon) = \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!} \left[\frac{d^n}{d\epsilon^n} f(x,X,t;\epsilon)\right]_{\epsilon=0} = \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!} f^{(n)}(y,Y,t) \tag{2.7}$$

where

$$f_n(x,X,t) = \left[\frac{\partial^n}{\partial\epsilon^n} f(x,X,t;\epsilon)\right]_{\epsilon=0} , \quad n \geq 0; \tag{2.8}$$

$$\frac{df}{d\epsilon}(x,X,t;\epsilon) = \frac{\partial f}{\partial\epsilon} + f_x.\frac{dx}{d\epsilon} + f_X.\frac{dX}{d\epsilon} , \tag{2.9a}$$

and

$$f^{(n)}(y,Y,t) = \left[\frac{d^n}{d\epsilon^n} f(x,X,t;\epsilon)\right]_{\epsilon=0} , \quad n \geq 0 . \tag{2.9b}$$

Notice that

$$f_0(x,X,t) = f(x,X,t;0) , \text{ and } f^{(0)}(y,Y,t) = f(y,Y,t;0) .$$

Using Eqs. (2.1a) and (2.1b), Eq. (2.9a) can be written as

$$\frac{df}{d\epsilon} = \frac{\partial f}{\partial\epsilon} + L_W f \tag{2.10}$$

where L_W is a linear operator called Lie derivative generated by W, and is defined by

$$L_W f = (f;W) = f_x.W_X - f_X.W_x \tag{2.11}$$

Taking $f = x,X,$ and F in Eq. (2.7), and using Eqs. (2.1a), (2.1b), and (2.1d), one obtains the following

$$x = y + \sum_{n=1}^{\infty} \frac{\epsilon^n}{n!} y^{(n)}(y,Y,t) \tag{2.12a}$$

$$X = Y + \sum_{n=1}^{\infty} \frac{\epsilon^n}{n!} Y^{(n)}(y,Y,t) \tag{2.12b}$$

$$H = K - \sum_{n=1}^{\infty} \frac{\epsilon^n}{n!} R^{(n)}(y,Y,t) \quad , \tag{2.12c}$$

where, for $n \geq 1$ we have

$$y^{(n)} = \left(\frac{d^{n-1}}{d\epsilon^{n-1}} W_X \right)_{\epsilon=0} \tag{2.13a}$$

$$Y^{(n)} = -\left(\frac{d^{n-1}}{d\epsilon^{n-1}} W_x \right)_{\epsilon=0} \tag{2.13b}$$

$$R^{(n)} = -\left(\frac{d^{n-1}}{d\epsilon^{n-1}} W_t \right)_{\epsilon=0} \quad . \tag{2.13c}$$

In particular, for a generating function W of the form

$$W(x,X,t;\epsilon) = \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!} W_{n+1}(x,X,t) \quad , \tag{2.14}$$

and $f(x,X,t;\epsilon)$ of the form given by Eq. (2.6), Eq. (2.10) yields

$$\frac{df}{d\epsilon} = \sum_{n \geq 0} \frac{\epsilon^n}{n!} f_n^{(1)}(x,X,t) \tag{2.15}$$

with

$$f_n^{(1)}(x,X,t) = f_{n+1} + \sum_{0 \leq m \leq n} C_m^n L_{m+1} f_{n-m} \tag{2.16}$$

where

$$C_m^n = \frac{n!}{(n-m)!m!} \quad , \tag{2.17a}$$

and

$$L_p f = (f;W_p) \ , \ p \geq 1 \quad . \tag{2.17b}$$

In general, for $k \geq 1$, $n \geq 0$, one obtains

$$\frac{d^k}{d\epsilon^k} f = \sum_{n\geq 0} \frac{\epsilon^n}{n!} f_n^{(k)}(x,X,t) \tag{2.18}$$

with

$$f_n^{(k)}(x,X,t) = f_{n+1}^{(k-1)} + \sum_{0\leq m\leq n} C_m^n \, L_{m+1} f_{n-m}^{(k-1)} \quad . \tag{2.19}$$

Now, letting $\epsilon = 0$ in the above equation we get the following. (For the remainder of this paper, this equation will be referred to as Deprit's equation.)

$$f_n^{(k)}(y,Y,t) = f_{n+1}^{(k-1)} + \sum_{0\leq m\leq n} C_m^n \, L_{m+1} \, f_{n-m}^{(k-1)} \tag{2.20}$$

where

$$L_p f = (f;W_p) = f_y . W_{pY} - f_Y . W_{py} \ , \ p \geq 1 \quad . \tag{2.21}$$

Notice that

$$f_n^{(0)}(y,Y,t) = f_n(y,Y,t), \text{ and } f_0^{(k)}(y,Y,t) = f^{(k)}(y,Y,t) \quad .$$

Deprit's equation, together with the functions $H^{(n)}$, $R^{(n)}$, $y^{(n)}$, and $Y^{(n)}$ can be best visualized from the triangles of Fig. 1 and Fig. 2.

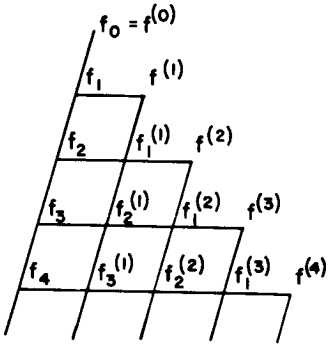


FIG. 1. RECURSIVE TRANSFORMATION OF AN ANALYTIC FUNCTION UNDER A LIE TRANSFORM.

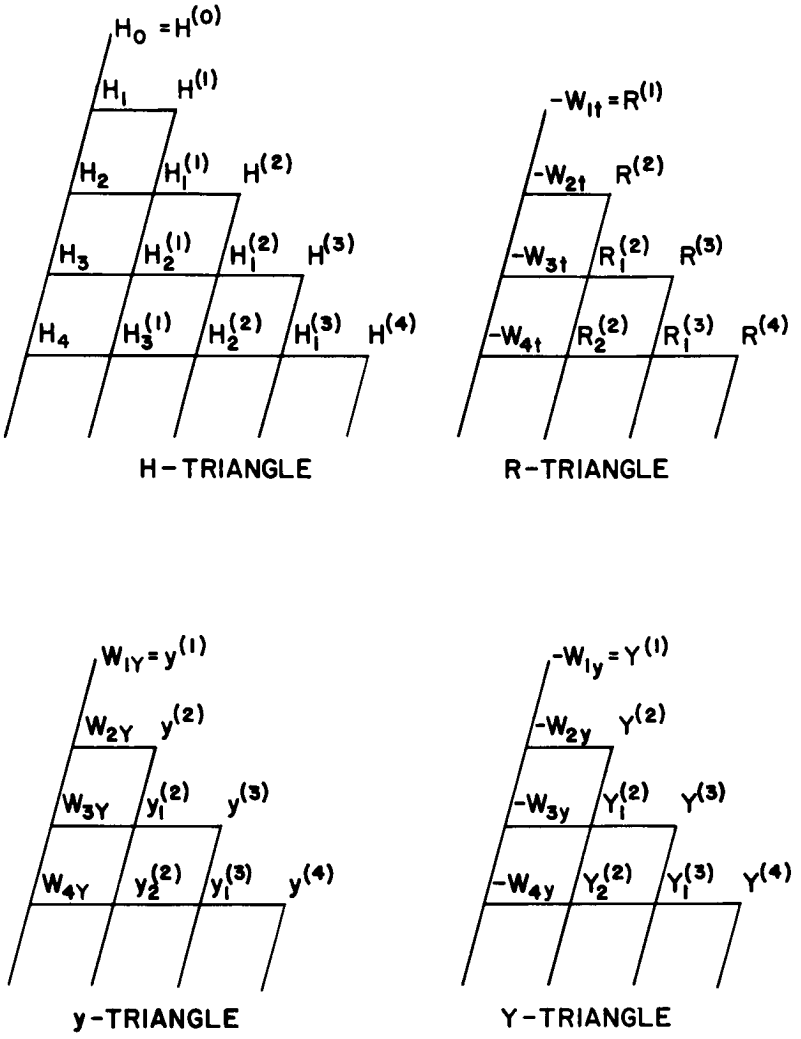


FIG. 2. TRIANGLES FOR THE HAMILTONIAN H, THE COORDINATES y, THE MOMENTA Y, AND THE REMAINDER R.

Finally, the inverse transformation can be written as

$$y = x + \sum_{n=1}^{\infty} \frac{\epsilon^n}{n!} x^{(n)}(x,X,t) \qquad (2.22a)$$

$$Y = X + \sum_{n=1}^{\infty} \frac{\epsilon^n}{n!} X^{(n)}(x,X,t) \quad . \qquad (2.22b)$$

To find the relation between the $x^{(n)}$'s and $y^{(n)}$'s, $X^{(n)}$'s and $Y^{(n)}$'s, one may eliminate x-y and X-Y between Eqs. (2.12a), (2.12b), and (2.22), and define the functions $q(x,X,t;\epsilon)$ and $Q(x,X,t;\epsilon)$ as follows:

$$q(x,X,t;\epsilon) = \sum_{n=1}^{\infty} \frac{\epsilon^n}{n!} x^{(n)}(x,X,t) = - \sum_{n=1}^{\infty} \frac{\epsilon^n}{n!} y^{(n)}(y,Y,t) \qquad (2.23a)$$

$$Q(x,X,t;\epsilon) = \sum_{n=1}^{\infty} \frac{\epsilon^n}{n!} X^{(n)}(x,X,t) = - \sum_{n=1}^{\infty} \frac{\epsilon^n}{n!} Y^{(n)}(y,Y,t) \qquad (2.23b)$$

Therefore, for $n \geq 1$ we have

$$q_n = x^{(n)}(x,X,t), \; q^{(n)} = - y^{(n)}(y,Y,t) \; , \qquad (2.24a)$$

$$Q_n = X^{(n)}(x,X,t), \; Q^{(n)} = - Y^{(n)}(y,Y,t) \; . \qquad (2.24b)$$

3. SIMPLIFIED GENERAL EXPANSIONS

Given the functions f_n, f_{n-1}, .., and f_0, Deprit (1968) constructed the required functions $f^{(n)}$, $f^{(n-1)}$, .., and $f^{(0)}$ by introducing the auxiliary functions $f_n^{(k)}$ and by moving recursively from the left diagonal of Fig. 1 towards the right diagonal. One might as well

construct the function $f^{(n)}$ only in terms of f_n and $f^{(n-1)}$, $f^{(n-2)}$ ..., $f^{(0)}$ or f_n in terms of $f^{(n)}$, $f^{(n-1)}$,..., and $f^{(0)}$ (which will be useful in the construction of the inverse transformation) by introducing a suitable linear operator. To show how this can be done, let us write Deprit's equation as

$$f_n^{(k)} = f_{n-1}^{(k+1)} - \sum_{m=0}^{n-1} C_m^{n-1} L_{m+1} f_{n-m-1}^{(k)}; \; n \geq 1, \; k \geq 0 \; . \qquad (3.1)$$

By successive elimination of the functions on the right hand side of the above equation one would eventually obtain $f_n^{(k)}$ in terms of $f^{(k+n)}$, $f^{(k+n-1)}$,..., $f^{(k)}$. Thus, one may assume the following form for $f_n^{(k)}$

$$f_n^{(k)} = f^{(k+n)} - \sum_{j=1}^{n} C_j^n G_j f^{(k+n-j)}; \; n \geq 1, \; k \geq 0 \; , \qquad (3.2)$$

where G_j is a linear operator and is a function of L_j, L_{j-1},..., and L_1. Substitution of Eq. (3.2) into Eq. (3.1) yields the following recursion relation for the linear operator G_j

$$G_j = L_j - \sum_{0 \leq m \leq j-2} C_m^{j-1} L_{m+1} G_{j-m-1}, \; 1 \leq j \leq n \; . \qquad (3.3)$$

For example

$$G_1 = L_1 \qquad (3.4a)$$

$$G_2 = L_2 - L_1 L_1 \qquad (3.4b)$$

$$G_3 = L_3 - L_1(L_2 - L_1 L_1) - 2L_2 L_1 \qquad (3.4c)$$

Using Eq. (3.2) with $f = y$ and Y, and taking $k = 1$, we obtain the following general recursive relations for $y^{(n)}$ and $Y^{(n)}$ of Eqs. (2.12a) and (2.12b)

$$y^{(n)} = W_{nY} + \sum_{1 \le j \le n-1} C_j^{n-1} G_j y^{(n-j)}, \quad n \ge 1 \quad , \tag{3.5a}$$

$$Y^{(n)} = - W_{ny} + \sum_{1 \le j \le n-1} C_j^{n-1} G_j Y^{(n-j)}, \quad n \ge 1 \quad . \tag{3.5b}$$

Using Eq. (3.2) with $f = q$ and Q of Eqs. (2.23a) and (2.23b), and taking $k = 0$, we obtain the following general formulae for $x^{(n)}$ and $X^{(n)}$

$$x^{(n)} = -y^{(n)} + \sum_{1 \le j \le n-1} C_j^n G_j y^{(n-j)}, \quad n \ge 1 \tag{3.6a}$$

$$X^{(n)} = -Y^{(n)} + \sum_{1 \le j \le n-1} C_j^n G_j Y^{(n-j)}, \quad n \ge 1 \tag{3.6b}$$

Now, $x^{(n)}(x,X,t)$ and $X^{(n)}(x,X,t)$ of Eqs. (2.22a) and (2.22b) are simply given by

$$x^{(n)}(x,X,t) = \left[x^{(n)}\right]_{\substack{y = x \\ Y = X}} \quad , \tag{3.7a}$$

$$X^{(n)}(x,X,t) = \left[X^{(n)}\right]_{\substack{y = x \\ Y = X}} \quad . \tag{3.7b}$$

Next, consider an indefinitely differentiable function $v(x,X,t)$ not explicitly dependent on ϵ. Using Eqs. (2.6), (2.7), and (3.2) with

$f_0 = v(x,X,t)$, $f_n = 0$ for $n \geq 1$, and $k = 0$, we obtain the following general formula

$$v(x,X,t) = \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!} v^{(n)}(y,Y,t) \tag{3.8a}$$

$$v^{(n)}(y,Y,t) = \sum_{1 \leq j \leq n} C_j^n G_j v^{(n-j)}(y,Y,t) \ , \ n \geq 1 \tag{3.8b}$$

where

$$v^{(0)}(y,Y,t) = v(y,Y,t) \ . \tag{3.8c}$$

Also the inverse relation can be written as

$$v(y,Y,t) = v(x,X,t) + \sum_{n=1}^{\infty} \frac{\epsilon^n}{n!} V^{(n)}(x,X,t) \ , \tag{3.9a}$$

elimination of $v(y,Y,t) - v(x,X,t)$ between Eqs. (3.8a) and (3.9a), and using Eqs. (2.6), (2.7), (3.2) with $k = 0$, and (3.8b) leads to

$$V^{(n)}(x,X,t) = - \left[G_n \, v(y,Y,t) \right]_{\substack{y = x \\ Y = X}} \ . \tag{3.9b}$$

Lastly, given the Hamiltonian $H(x,X,t;\epsilon)$ in the form

$$H(x,X,t;\epsilon) = \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!} H_n(x,X,t) \tag{3.10}$$

one can construct the transformed Hamiltonian $K(y,Y,t;\epsilon)$ in the form

$$K(y,Y,t;\epsilon) = \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!} K_n(y,Y,t) \tag{3.11}$$

and this can be done as follows. Using Eq. (2.7), H can be written as

$$H(x,X,t;\epsilon) = \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!} H^{(n)}(y,Y,t) \quad . \tag{3.12}$$

Combination of Eq. (3.12) with Eq. (2.12c) and Eq. (3.11) yields

$$K_0 = H_0 \tag{3.13a}$$

$$K_n = H^{(n)} + R^{(n)} \quad , \quad n \geq 1 \quad . \tag{3.13b}$$

Setting $k = 1$ and $f = H + R$ in Eq. (3.2) leads to

$$H_n^{(1)} + R_n^{(1)} = K_{n+1} - \sum_{j=1}^{n} C_j^n \, G_j \, K_{n-j+1} \quad , \quad n \geq 1 \quad . \tag{3.14}$$

But from the R and H triangles of Fig. 2, we have

$$R_n^{(1)} = - \frac{\partial W_{n+1}}{\partial t} \quad , \quad n \geq 0 \tag{3.15}$$

$$H_n^{(1)} = H_{n+1} + \sum_{m=0}^{n} C_m^n \, L_{m+1} \, H_{n-m} \quad , \quad n \geq 0 \quad . \tag{3.16}$$

Therefore, the simplified general recursive relation of the transformed Hamiltonian is given by

$$K_0 = H_0 \tag{3.17a}$$

$$K_n = H_n + \sum_{1 \leq j \leq n-1} \left(C_{j-1}^{n-1} \, L_j \, H_{n-j} + C_j^{n-1} \, G_j \, K_{n-j} \right) - \frac{DW_n}{Dt} \quad , \quad n \geq 1 \tag{3.17b}$$

where

$$\frac{DW_n}{Dt} = \frac{\partial W_n}{\partial t} - L_n H_0 \ , \ n \geq 1 \quad . \tag{3.18}$$

For example

$$K_1 = H_1 - DW_1/Dt \tag{3.19a}$$

$$K_2 = H_2 + L_1 H_1 + G_1 K_1 - DW_2/Dt \tag{3.19b}$$

$$K_3 = H_3 + L_1 H_2 + 2L_2 H_1 + 2G_1 K_2 + G_2 K_1 - DW_3/Dt \tag{3.19c}$$

$$K_4 = H_4 + L_1 H_3 + 3L_2 H_2 + 3L_3 H_1 + 3G_1 K_3 + 3G_2 K_2 + G_3 K_1 - DW_4/Dt \quad . \tag{3.19d}$$

the operators G_1, G_2, and G_3 being as defined for Eqs. (3.4a) to (3.4c).

4. MODIFIED GENERAL EXPANSIONS

In the simplified formulae obtained in Section 3, the rate of increase of the number of the Poisson brackets with respect to the order of perturbation can be reduced if one uses intermediate functions like $f_{j,n} = G_j f_n$ or $G_j f^{(n)}$ to be saved for later use in computation. Thus, this leads to the following recursive relationships:

a) For $y^{(n)}$, $x^{(n)}$, $Y^{(n)}$, and $X^{(n)}$ of Eqs. (3.5) and (3.6)

$$y^{(n)} = W_{ny} + \sum_{1 \leq j \leq n-1} C_j^{n-1} \, y_{j,n-j} \tag{4.1a}$$

$$x^{(n)} = -y^{(n)} + \sum_{1 \leq j \leq n-1} C_j^{n} \, y_{j,n-j} \tag{4.1b}$$

$$Y^{(n)} = -W_{ny} + \sum_{1\le j\le n-1} C_j^{n-1}\, Y_{j,n-j} \tag{4.1c}$$

$$X^{(n)} = -Y^{(n)} + \sum_{1\le j\le n-1} C_j^{n}\, Y_{j,n-j} \tag{4.1d}$$

where

$$y_{j,i} = L_j\, y^{(i)} - \sum_{0\le m\le j-2} C_m^{j-1}\, L_{m+1}\, y_{j-m-1,i} \tag{4.2a}$$

$$Y_{j,i} = L_j\, Y^{(i)} - \sum_{0\le m\le j-2} C_m^{j-1}\, L_{m+1}\, Y_{j-m-1,i}\ ; \tag{4.2b}$$

(b) For $v^{(n)}$ and $V^{(n)}$ of Eqs. (3.8b) and (3.9b)

$$v^{(n)} = \sum_{1\le j\le n} C_j^{n}\, v_{j,n-j} \tag{4.3a}$$

$$V^{(n)}(x,X,t) = -\left[v_{n,0}\right]_{\substack{y=x\\ Y=X}} \tag{4.3b}$$

$$v_{j,i} = L_j\, v^{(i)} - \sum_{0\le m\le j-2} C_m^{j-1}\, L_{m+1}\, v_{j-m-1,i}\ ; \tag{4.4}$$

(c) For K_n of Eq. (3.17b)

$$K_n = H_n + \sum_{1\le j\le n-1} \left(C_{j-1}^{n-1}\, L_j H_{n-j} + C_j^{n-1}\, K_{j,n-j} \right) - \frac{DW_n}{Dt}\ ,\ n \ge 1 \tag{4.5a}$$

where

$$K_{j,i} = L_j K_i - \sum_{0\le m\le j-2} C_m^{j-1}\, L_{m+1}\, K_{j-m-1,i}\ . \tag{4.5b}$$

Notice that for $K_i = 0$, $K_{j,i} = 0$ for all j's. For example,

$$K_2 = H_2 + L_1H_1 + K_{1,1} - DW_2/Dt \tag{4.6a}$$

$$K_3 = H_3 + L_1H_2 + 2L_2H_1 + 2K_{1,2} + K_{2,1} - DW_3/Dt \tag{4.6b}$$

$$K_4 = H_4 + L_1H_3 + 3L_2H_2 + 3L_3H_1 + 3K_{1,3} + 3K_{2,2} + K_{3,1} - DW_4/Dt \tag{4.6c}$$

where

$$K_{1,1} = L_1K_1 \tag{4.7a}$$

$$K_{1,2} = L_1K_2, \; K_{2,1} = L_2K_1 - L_1K_{1,1} \tag{4.7b}$$

$$\left.\begin{aligned} K_{1,3} &= L_1K_3, \; K_{2,2} = L_2K_2 - L_1K_{1,2}, \\ K_{3,1} &= L_3K_1 - L_1K_{2,1} - 2L_2K_{1,1} \; . \end{aligned}\right\} \tag{4.7c}$$

The construction of the transformed Hamiltonian using the scheme presented by Eqs. (4.5a), (4.5b), and (3.18) is simpler and requires less computer time and storage than the scheme presented by Deprit (1968). A considerable amount of this reduction is due to the fact that the sums $H^{(n)} + R^{(n)}$ as well as $H_n^{(1)} + R_n^{(1)}$ in Eqs. (3.13b) and (3.14) were considered as single quantities as if the transformed Hamiltonian was constructed from a single triangle whose end products are $H^{(n)} + R^{(n)}$ and whose starting elements are $H_n^{(1)} + R_n^{(1)}$. Further reduction in the computer requirements can be achieved if some of the K_i's vanish, in which case $K_{j,i}$ also vanishes for all possible values of j.

Equations (3.17) or (4.5) and (3.18) are directly applicable to nonlinear resonant problems in which H_0 is a function of only the

action variables X, while $H_n (n \geq 1)$ depend trigonometrically on the angle variables x and possibly the time t. It is desirable to transform to new variables so that the resulting Hamiltonian contains, together with the new action variables Y, only certain slowly-varying "long-period" combinations of the new angle variables y and the time t. Equation (3.17) or (4.5) may be used to define the W_n's successively so as to remove all "short-period" terms from the K_n's; such a W_n is unique up to an arbitrary additive long-period term.

The equations obtained are now being used in a fourth-order analysis of the motion (stability and periodic orbits) of a particle in the neighborhood of L_4 of the earth-moon system in the presence of the sun. In this problem, the following parameters are treated as first order small quantities: distance from L_4/earth-moon distance, eccentricity of the moon's orbit around the earth, moon mass/earth mass, mean motion of the sun/mean motion of the moon. The additional parameter (earth-moon distance/earth-sun distance) is treated as a second order small quantity.

5. ACKNOWLEDGEMENTS

The author gratefully thanks his advisor, Professor John V. Breakwell, for his helpful suggestions and assistance. The suggestions and comments of Dr. Andre´ Deprit, Dr. Ali Nayfeh, and Dr. Jacques Henrard are appreciated.

This research was supported by the National Aeronautics and Space Administration, under Contract No. NsG 133-61.

REFERENCES

Breakwell, J. V., and R. Pringle, Jr., 1966, Progress in Astronautics, Vol. 17, Academic Press, Inc., New York, pp. 55-73.

Brouwer, D. and Clemence, G. M., 1961, Methods of Celestial Mechanics, Academic Press, New York, New York.

Deprit, A., 1966, Proceedings of I.A.U., Symposium No. 25, Academic Press, pp. 170-175.

Deprit, A., Henrard , J., and Rom, A. R. M., 1967, Icarus 6, 381-406.

Deprit, A., 1968, "Canonical Transformations Depending on a small Parameter", Boeing Scientific Research Labs., The Boeing Company, Seattle, Washington, document No. D1-82-0755.

Schechter, H. B., 1968, AIAA Journal Vol. 6, No. 7.

THE FORMAL SOLUTION OF THE n-BODY PROBLEM

By P. Sconzo

and

D. Valenzuela
International Business Machines
Cambridge Advanced Space Systems Department
Cambridge, Massachusetts

P. Sconzo

D. Valenzuela

THE FORMAL SOLUTION OF THE n-BODY PROBLEM

ABSTRACT: The power series solution of the equations of motion of a system of n point-masses is presented. This solution is a formal one in the time domain. The origin of the series expansions is a non-collision point. A procedure has been developed using three fundamental recursion formulas, one of which involves a special differential operator. Some of these analytical formulations have been programmed in the PL/I FORMAC language. Results are presented.

Both authors are located at the IBM Cambridge Advanced Space Systems Department, FSD, Cambridge, Massachusetts.

THE FORMAL SOLUTION OF THE n-BODY PROBLEM

1. Recent applications of the methods of celestial mechanics to problems of space flights impose severe requirements upon the quality of the solution. Quality stands here for high level of accuracy in the computed position of a space probe when its motion takes place under the perturbations exerted by many bodies. Means to satisfy those requirements are offered by well-known numerical integration techniques which can be applied to the equations of motion. Although efficient from a computational point of view, these techniques are regarded in general as being a rather crude approach to the solution of the problem. Besides, it might be desirable, from a theoretical point of view, that the solution be obtained in an analytical form, for instance that it be constructed as a time-power series. In this paper we show how this analytical goal can be achieved. We will give a recursive method to construct the terms of these series up to any desired high-order power of the independent variable (t). The formal solution thus obtained could be used to cover an arc of the trajectory much longer than the step used in any numerical integration procedure. This solution is valid, of course, in a neighborhood of t = 0, origin of the series expansion, which is assumed not to be a collision point.

   The crucial problem to be solved is to construct the series expansion of the inverse cube of the distance between bodies. This is achieved introducing some auxiliary functions and operating on them with an "ad-hoc" differential operator. The end result provides the coefficients of the series as functions of the initial conditions which must be satisfied by the equations of motion.

2. We consider the motion of n bodies in a Newtonian potential field. Let $m_i (i = 1,2,3,...n)$ be the masses of these bodies, to be considered as point-masses, and x_i, y_i, z_i their Cartesian coordinates referred to an inertial reference frame. For the sake of generality, we suppose that none of the masses is negligible and that none of the bodies is constrained to move along a prescribed trajectory.

THE FORMAL SOLUTION OF THE n-BODY PROBLEM

The equations of motion can be written as follows

$$D^{(2)}x_i = k\sum_{j=1}^{n} m_j \mu_{ij} a_{ij} \qquad (j \neq i) \tag{1}$$

where

$$\mu_{ij} = r^{-3}_{ij} \tag{2}$$

$$r^2_{ij} = a^2_{ij} + b^2_{ij} + c^2_{ij} \tag{3}$$

$$a_{ij} = x_j - x_i,\ b_{ij} = y_j - y_i,\ c_{ij} = z_j - z_i, \tag{4}$$

and similarly for y_i and z_i. The symbol D stands for $\frac{d}{dt}$, and k is the gravitational constant. If the fundamental units of length, mass and time are appropriately chosen, then we can take k = 1.

The set of equations (1) constitutes a system of 3n differential equations of the second order.

Let now

$$x_{i0} = x_i(0),\ x_{i1} = (Dx_i)_0 \tag{5}$$

be the initial conditions. The formal solution of (1) as a Taylor series in t is

$$x_i = \sum_{\nu=0}^{\infty} x_{i\nu} t^{\nu} \tag{6}$$

where

$$x_{i\nu} = \frac{1}{\nu!}(D^{(\nu)}x_i)_0$$

The first two coefficients in (6) are given by (5). The third coefficient x_{i2} is given by the right-hand side of equation (1), evaluated

at t = 0 and multiplied by one-half. Then, the successive coefficients can be constructed by an iterative procedure. In fact, setting

$$\mu_{ij} = \sum_{\nu=0}^{\infty} \mu_{ij\nu} t^{\nu} \tag{7}$$

we get from (1) by a well-known procedure

$$x_{i,\nu+2} = \frac{1}{(\nu+2)(\nu+1)} \sum_{j=1}^{n} m_j \sum_{p=0}^{\nu} \mu_{ijp} a_{ij\nu-p}, \qquad (\nu \geq 0). \tag{8}$$

This equation can be used recursively after we have learned how to compute $\mu_{ij1}, \mu_{ij2}, \ldots, \mu_{ij\nu}$ in terms of the initial conditions (5). We put emphasis on the fact that there are $\frac{1}{2}n(n-1)$ functions μ_{ij} which should be handled simultaneously due to the coupling of the subscripts i,j. This implies that the algebraic manipulations to be performed will be very lengthy, even for relatively small values of n. We will, however, omit from now on the subscripts i,j. The notation used by Sconzo [1] in his investigation on the tridimensional non-restricted three-body problem will also be used here.

3. We introduce the first of our auxiliary functions by the definition

$$s = aDa + bDb + cDc \tag{9}$$

For compactness we rewrite s as follows

$$s = S[aDa] \tag{10}$$

where the symbol S[...] has the meaning of a sum extended over terms in b and c similar to that in a.

Then, successively differentiating the function μ it is not difficult to recognize that

$$\mu_{\nu} = \frac{1}{\nu!} D^{(\nu)} \mu = -\frac{3\mu}{\nu!} (\rho P_{\nu} + Q_{\nu}) \tag{11}$$

where P_{ν} is an expression in σ, ε and the derivatives of s of order

higher than the second, and Q_ν a polynomial in σ,ε where ρ, σ, and ε are new auxiliary functions defined as follows

$$\rho = r^{-2},\quad \sigma = \rho s,\quad \varepsilon = \rho D s \tag{12}$$

In order to prove (11) one has to observe that the first derivatives of all the auxiliary functions so far introduced can be expressed in terms of the functions themselves. In fact, it is

$$D\mu = -3\mu\sigma,\quad D\rho = -2\rho\sigma,\quad D\sigma = \varepsilon - 2\sigma^2 \tag{13}$$

$$D\varepsilon = -2\sigma\varepsilon + \rho D^{(2)}s \ * \tag{14}$$

In our recursive procedure the equations (8) and (11) are pivotal formulas, together with the formula obtained by applying Leibnitz rule to the right-hand side of equation (10). Distinguishing the cases where the order of differentiation is odd or even, the latter formula is, respectively,

$$D^{(2\nu+1)}s = S\left[\frac{1}{2}\binom{2\nu+2}{\nu+1}\{D^{(\nu+1)}a\}^2 + \sum_{p=1}^{\nu+1}\binom{2\nu+2}{\nu+1+p}D^{(\nu+1+p)}aD^{(\nu+1-p)}a\right] \tag{15a}$$

$$D^{(2\nu)}s = S\left[\sum_{p=1}^{\nu+1}\binom{2\nu+1}{\nu+p}D^{(\nu+p)}aD^{(\nu+1-p)}a\right]. \tag{15b}$$

We notice in passing that the functions μ, ρ, s, σ and ε can be explicitly expressed in terms of the initial conditions (5). The derivatives of s of order greater than the second become instead implicitly defined in terms of the same initial conditions by virtue of (15a) and (15b).

---

*We notice the analogy with the two body problem formulation[2]. In this particular case there is only one function μ and the auxiliary function s satisfies the differential equation

$$D^{(2)}s = -\mu s$$

Thus, equation (14) reduces to

$$D\varepsilon = -\sigma(2\varepsilon + \mu)$$

and the whole procedure is greatly simplified.

4. The method we have described can be considered completed if we succeed in giving explicit expressions for P_ν and Q_ν for any desired value of ν.

To this end we consider the following operator

(16) $$\Theta = D - 5\sigma$$

Then, a simple algebraic manipulation provides

(17) $$P_{\nu+1} = \Theta P_\nu + A_\nu$$

$$Q_{\nu+1} = -3\sigma Q_\nu + B_\nu$$

where

(18) $$A_\nu = \frac{\delta Q_\nu}{\delta \varepsilon} \cdot D^{(2)}{}_s$$

$$B_\nu = \frac{\delta Q_\nu}{\delta \sigma}(\varepsilon - 2\sigma^2) - 2\sigma\varepsilon\frac{\delta Q_\nu}{\delta \varepsilon}$$

Thus, starting from

$$P_0 = 0, \; Q_0 = -\frac{1}{3}$$

any expression can be generated, by hand for lower indexed, by a computer for higher indexed functions P_ν and Q_ν. A program written in the PL/I FORMAC language has generated these functions, consequently the derivatives of μ, up to orders far exceeding any practical need. We list in the table appended below the first six of these functions. For $\nu > 6$, the expressions become very lengthy, and this is the only reason which prevents their presentation in this paper.

The problem of finding the formal solution (6) of the equations of type (1) can thus be considered solved.

TABLE OF THE FUNCTIONS $\frac{1}{\nu!}P_\nu$ AND $\frac{1}{\nu!}Q_\nu$

$$P_1 = 0 \qquad Q_1 = \sigma$$

$$\frac{1}{2!}P_2 = 0 \qquad \frac{1}{2!}Q_2 = \frac{1}{2}(\varepsilon - 5\sigma^2)$$

$$\frac{1}{3!}P_3 = \frac{1}{6}D^{(2)}_s \qquad \frac{1}{3!}Q_3 = \frac{5}{6}\sigma(-3\varepsilon + 7\sigma^2)$$

$$\frac{1}{4!}P_4 = \frac{1}{24}D^{(3)}_s - \frac{5}{6}\sigma D^{(2)}_s \qquad \frac{1}{4!}Q_4 = -\frac{5}{8}\varepsilon^2 + \frac{35}{8}\sigma^2(2\varepsilon - 3\sigma^2)$$

$$\frac{1}{5!}P_5 = \frac{1}{120}D^{(4)}_s - \frac{5}{24}\sigma D^{(3)}_s - \frac{5}{12}(\varepsilon - 7\sigma^2) D^{(2)}_s \qquad \frac{1}{5!}Q_5 = \frac{7}{8}\sigma(5\varepsilon^2 - 30\sigma^2\varepsilon + 33\sigma^4)$$

$$\frac{1}{6!}P_6 = \frac{1}{720}D^{(5)}_s - \frac{1}{24}\sigma D^{(4)}_s - \frac{5}{48}(\varepsilon - 7\sigma^2) D^{(3)}_s + \frac{35}{12}\sigma(\varepsilon - 3\sigma^2) D^{(2)}_s - \frac{5}{72}\rho (D^{(2)}_s)^2$$

$$\frac{1}{6!}Q_6 = \frac{35}{48}\varepsilon^3 - \frac{7}{16}\sigma^2(45\varepsilon^2 - 165\sigma^2\varepsilon + 143\sigma^4)$$

REFERENCES

[1] P. Sconzo, NASA-ERC Publication SP-141, 1967 and Astronomische Nachrichten, vol. 290, p. 163, 1967.

[2] P. Sconzo, A. R. LeSchack, R. Tobey, Astr. Journal, vol. 70, p. 269, 1965, and P. Sconzo, D. Valenzuela, NASA-ERC, Publication PM-67-21, 1968.

THE LONG PERIOD BEHAVIOR OF A CLOSE LUNAR ORBITER INCLUDING THE INDIRECT SOLAR GRAVITY PERTURBATION

by

Robert Dasenbrock

Doctoral Candidate and Research Assistant
Department of Aeronautics and Astronautics

STANFORD UNIVERSITY
Stanford Electronics Laboratory
Stanford, California 94305

Advisor: Professor John V. Breakwell

THE LONG PERIOD BEHAVIOR OF A CLOSE LUNAR ORBITER INCLUDING THE INDIRECT SOLAR GRAVITY PERTURBATION

by

Robert Dasenbrock*

ABSTRACT

The long period behavior of a lunar orbiter is considered. Of special interest are the effects due to the inclination of the apparent Earth's orbit about the Moon and those effects described by the laws of Cassini on the equations of motion. The first part of the paper is restricted to low orbits where the lunar gravity field dominates the terrestrial perturbation and to higher orbits of low inclinations where the argument of pericenter circulates through an angle of 360 degrees. The last part of the paper deals with near polar orbits where the indirect solar perturbation as described by the laws of Cassini is most important. Long-term stable positions for the orbit plane are found.

---

* Doctoral Candidate and Research Assistant, Department of Aeronautics and Astronautics, Stanford University, Stanford, California (January 1969)

CLOSE LUNAR ORBITER LONG PERIOD BEHAVIOR

NOMENCLATURE

a	semi-major axis
e	eccentricity
e_E	eccentricity of apparent Earth's orbit about the Moon
F	the negative of the Hamiltonian
F_0, F_1, F_2, F_3	components of F
g	argument of perigee
G	$\sqrt{\mu a(1-e^2)}$, canonically conjugate to g
$\vec{h}$	angular momentum vector of satellite
h	position of the ascending node
H	$\sqrt{\mu a(1-e^2)}\cos i$, canonically conjugate to h
$\mathcal{H}$	Hamiltonian
$\mathcal{H}_r$	component of the Hamiltonian
i	inclination
I_E	inclination of the Earth's orbit to the lunar equator = 6 degrees 44 min
J_{20}, J_{22}, J_3, J_4	lunar gravity coefficients
ℓ	mean anomaly
L	$\sqrt{\mu a}$, canonically conjugate to ℓ
$\mathfrak{L}$	Lagrangian
p	$a(1-e^2)$
$\vec{p}_I$	momentum canonically conjugate to coordinate, $\vec{r}_I$, in inertial space
$\vec{p}_r$	momentum canonically conjugate to coordinate, $\vec{r}_r$, in the rotating frame of reference
$\vec{r}_I$	coordinate in inertial space
r_p	perigee heigth = a(1-e)
$\vec{r}_r$	coordinate in the rotating frame of reference
R_m	lunar radius

S	generating function
S_0, S_1, S_2	components of S
$V(r)$	arbitrary potential function
θ_E	Earth coordinate
Θ	momentum canonically conjugate to θ_E
μ	gravitational constant of the Moon
ω	argument of pericenter = g
Ω	position of the ascending node = h

CLOSE LUNAR ORBITER LONG PERIOD BEHAVIOR

I. INTRODUCTION

Some attention in recent years has been focused on the problem of determining the motion of a lunar satellite. The problem is complicated by the peculiar nature of the Moon's gravity field. Early attempts by several authors[1,2,3] on the solution of this problem were made by assuming the Moon to be in nearly hydrostatic equilibrium. Thus only the J_{20} and J_{22} gravity coefficients were carried in the equations of motion. The higher harmonics J_3, J_4, etc., were either ignored or assumed to be of order J_{20}^2. Independent determinations of the gravity coefficients by both the U.S.[4] and the U.S.S.R.[5] invalidate this assumption. It appears from the early data that the oblateness coefficient, J_{20}, has a value of approximately -2.0×10^{-4}. However preliminary data from Lunar Orbiters I through V still gives no conclusive evidence on the absolute values for the higher gravity coefficients. It appears at this time that these are all at most of order 10^{-5}.

The lunar orbiter problem is further complicated by the large perturbation caused by the Earth. For an orbiter of moderate height, say 800 to 2000 km above the surface, the terrestrial perturbation is roughly equal to the oblateness effect of the Moon's gravity field.

Of primary interest will be the long period effects, i.e., those fluctuations in the orbital elements having periods of several months or longer. Short period variations, all of which have much smaller amplitudes, will be averaged out. For a discussion of these latter effects the reader is referred to papers by Giacaglia[2] and Osterwinter[3]. The lunar gravity coefficients, J_{20}, J_{22}, J_3, J_4 will be retained in the equations of motion. Cassini's laws on the figure of the Moon will be considered in their classical form, i.e., the smaller effects of the physical librations will be ignored.

CLOSE LUNAR ORBITER LONG PERIOD BEHAVIOR

The relative effect of the eccentricity of the terrestrial path on the short period variations is of order e_E (=0.055). However its effect on the long period behavior is of order e_E^2 ($=0.055^2$) and thus will be neglected. Such is also the case with terms involving $e_E \sin I_E$. The small effects of the solar radiation pressure along with the direct solar gravity effect will not be considered.

II. CHOICE OF REFERENCE FRAME

As the behavior of close lunar orbiters is of primary interest, a reference frame coincident with the lunar equator is most convenient. This is especially so if the higher harmonics of the Moon's gravity field are considered. However, as a consequence of Cassini's laws, the plane of the lunar equator is not fixed in space. Cassini's laws state that the plane of the lunar equator, the ecliptic, and the plane of the Moon's orbit all coincide in a common line (ignoring physical librations). This line is the node of the lunar orbit as referenced to the ecliptic. It is convenient to choose this line (the ascending node) as the x-axis of the reference frame. The lunar axis of rotation is the z-axis. Thus full advantage is taken of the geometry of the system. This is described in Fig. (1). The system rotates in retrograde manner with a period of about 18.5 years.

If one is to work in the rotating system just described, the equations of motion, derived for a satellite moving in an inertial frame, must be modified. It is suggested that this modification take the form of an an additional perturbing term in the Hamiltonian. The system of reference is rotating with angular velocity components

$$\begin{aligned} \omega_x &= 0.0 && \text{rad/sec} \\ \omega_y &= 3.0\times10^{-10} && \text{rad/sec} \\ \omega_z &= -1.07\times10^{-8} && \text{rad/sec} \end{aligned} \qquad (1)$$

CLOSE LUNAR ORBITER LONG PERIOD BEHAVIOR

In an inertial frame of reference the Lagrangian is

$$\mathfrak{L} = \dot{\vec{r}}_I^{\,2}/2 - V(r_I) \tag{2}$$

In the rotating frame

$$\mathfrak{L} = (\dot{\vec{r}}_r + \vec{\omega}\times\vec{r}_r)^2/2 - V(r_r) \tag{3}$$

$$\vec{p}_r = \frac{\partial \mathfrak{L}}{\partial \dot{\vec{r}}_r} = \dot{\vec{r}}_r + \vec{\omega}\times\vec{r}_r \tag{4}$$

The Hamiltonian is

$$\mathcal{H} = \vec{p}_r\cdot\dot{\vec{r}}_r - \mathfrak{L} \tag{5}$$

Expressing $\mathcal{H}$ in terms of $(\vec{p}_r, \vec{r}_r)$ one obtains

$$\mathcal{H} = \frac{1}{2}\,\vec{p}_r\cdot\vec{p}_r + V(r_r) - \vec{\omega}\cdot\vec{h} \tag{6}$$

where $\vec{h}$ is the angular momentum of the satellite whose components are

$$\begin{aligned} h_x &= G \sin i \sin h \\ h_y &= -G \sin i \cos h \\ h_z &= G \cos i = H \end{aligned} \tag{7}$$

where G,H,h are the usual Delaunay variables. The additional term to be added to the Hamiltonian is

$$\begin{aligned} \mathcal{H}_r &= -\vec{\omega}\cdot\vec{h} = -\omega_y h_y - \omega_z h_z \\ \mathcal{H}_r &= +\omega_y G \sin i \cos h - \omega_z H \end{aligned} \tag{8}$$

CLOSE LUNAR ORBITER LONG PERIOD BEHAVIOR

III. THE DISTURBING FUNCTION

The computation of the disturbing function due to the perturbing effects of the Earth and Moon is straightfoward but lengthy and thus will not be reproduced here. Employing a result due to Brouwer[7] and Kozai , the Hamiltonian in mixed Keplerian and Delaunay variables is

$$
\begin{aligned}
-\mathcal{H} = F = {} & \frac{\mu^2}{2L^2} - n_E \Theta - \omega_y G \sin i \cos h + \omega_z H \\
& + \frac{1}{4}\frac{\mu^4 R_m^2}{L^3 G^3}\left\{J_{20}(1-3\cos^2 i) + 6J_{22}\sin^2 i \cos 2(h-\theta_E)\right\} \\
& - \frac{3}{8}\frac{\mu^5 R_m^3}{L^3 G^5}\left\{J_3 e \sin i\,(1-5\cos^2 i)\sin g\right\} \\
& - \frac{3}{128}\frac{\mu^6 R_m^4}{L^3 G^7} J_4\left\{(3-30\cos^2 i + 35\cos^4 i)(2+3e^2)\right. \\
& \qquad \left. -10\, e^2 \cos 2g(1-8\cos^2 i + 7\cos^4 i)\right\} \\
& + \frac{n_E^2 a^2}{16}\left[(2+3e^2)\left\{3\cos^2 i - 1 + 3\sin^2 i \cos 2(h-\theta_E)\right\}\right. \\
& \qquad + 15e^2\left\{\frac{1}{2}(1+\cos i)^2 \cos 2(g+h-\theta_E) + \sin^2 i \cos 2g\right. \\
& \qquad \left. + \frac{1}{2}(1-\cos i)^2 \cos 2(g-h+\theta_E)\right\} \\
& + \sin I_E \left\{(6+9e^2)\sin 2i \cos h - \cos(h-2\theta_E)\right. \\
& \qquad -30e^2 \sin i \cos i \cos h \cos 2g \\
& \qquad -\sin h \sin 2g + \sin(h-2g)\sin 2g \\
& \qquad \left. -\cos i \cos(h-2\theta_E)\cos 2g\right\} \\
& + \sin^2 I_E \left\{(6+9e^2)(\sin^2 i - \sin^2 h - \cos^2 h \cos^2 i)\right. \\
& \qquad \left. (\sin^2 h + \cos^2 h \cos^2 i + \sin^2 i)\cos 2\theta_E\right.
\end{aligned}
$$

$$+30e^2\left\{\cos i \sin h \cos h \cos 2\theta_E - \cos i \sin h \cos h \right\}\sin 2g$$
$$-15e^2\left\{\sin^2 h + \sin^2 i - \cos^2 h \cos^2 i + (\cos^2 h \cos^2 i - \sin^2 h - \sin^2 i)\cos 2\theta_E\right\}\cos 2g \Big\}\Big] \qquad (9)$$

The short period terms containing ℓ, the mean anomaly, have been averaged out. Θ is canonically conjugate to θ_E, the Earth coordinate, and F is the negative of the Hamiltonian. This convention will be used throughout.

It is desireable to write the Hamiltonian in the form

$$F = F_0 + F_1 + F_2 + \ldots$$

where F_0 is of order unity. F_1 is of order 10^{-2}, F_2 is of order 10^{-4} and so on. To determine the order of each term in Eq.(9), Fig.(2) is found useful. The terms $\mu^2/2L^2$ and $n_E\Theta$ are seen to belong to F_0 and F_1 respectively. Terms assigned to F_2 are $\omega_z H$ and the contribution associated with J_{20} and J_{22}. The Earth perturbation and those effects due to J_3 and J_4 belong to F_3. The disturbing function is thus of the form

$$F(L,G,H,\Theta,g,h,\theta_E) = F_0(L) + F_1(\Theta) + F_2(L,G,H,g,h,\theta_E) + F_3(L,G,H,g,h,\theta_E) + \ldots$$

At this point one wishes to eliminate all terms in F containing θ_E. This is accomplished by means of a stationary generating function

$$S(L',G',H',\Theta',g,h,\theta_E) = L'\ell + G'g + H'h + \Theta'\theta_E + S_1(L',G',H',g,h,\theta_E) + S_2(L',G',H',g,h,\theta_E) + \ldots \qquad (10)$$

such that the new Hamiltonian $F^*(L',G',H',\Theta',g',h')$ does not contain θ_E. The new coordinates are related to the old by the relations

$$L = \frac{\partial S}{\partial \ell} = L' \qquad \ell' = \frac{\partial S}{\partial L'} = \ell + \frac{\partial S_1}{\partial L'} + \frac{\partial S_2}{\partial L'} + \ldots \tag{11}$$

with similar relations for the other variables. From the relation

$$F^*(L',G',H',\Theta',g',h',\Theta_E') = F(L,G,H,\Theta,\ell,g,h,\Theta_E)$$

the following equations are derived for S_1 and S_2.

$$F_2^*(L',G',H',g,h,-) = F_2(L',G',H',g,h,\Theta_E) - n_E \frac{\partial S_1}{\partial \Theta_E} \tag{12}$$

$$F_3^* + \frac{\partial F_2^*}{\partial g}\frac{\partial S_1}{\partial G'} + \frac{\partial F_2^*}{\partial h}\frac{\partial S_1}{\partial H'} = F_3 + \frac{\partial F_2}{\partial G'}\frac{\partial S_1}{\partial g} + \frac{\partial F_2}{\partial H'}\frac{\partial S_1}{\partial h} - n_E \frac{\partial S_2}{\partial \Theta_E} \tag{13}$$

choosing $\frac{\partial S_1}{\partial \Theta_E}$ and $\frac{\partial S_2}{\partial \Theta_E}$ to cancel the periodic parts of Eqs. (12) and (13) respectively, the new Hamiltonian is

$$F^* = \frac{\mu^2}{2L'^2} - n_E\Theta' + F_2^*(L',G',H',g',h') + \overline{\frac{\partial F_2}{\partial G'}\frac{\partial S_1}{\partial g}} + \overline{\frac{\partial F_2}{\partial H'}\frac{\partial S_1}{\partial h}} + F_3^*(L',G',H',g',h') \tag{14}$$

As F^* does not contain Θ_E', Θ' is constant and will be dropped in the following discussion.

In order to use the von Zeipel method to determine the long period behavior of the orbital elements, F must be of the form

$$F^* = \frac{\mu^2}{2L'^2} + F_2^*(L',G',H') + F_3^*(L',G',H',g',h') \tag{15}$$

where F_3^* is dominated by F_2^*. This can be done if the orbits under consideration are restricted. The lower orbits, where J_{20} is the dominant perturbation, automatically fall into this category. Also included are the higher orbits provided the inclination remains low.

F_2^* in Eq. (15) is in mixed variables

$$F_2^* = \frac{1}{4}\frac{\mu^4 R_m^2}{L'^3 G'^3} J_{20}(1-3\cos^2 i) - \frac{n_E^2 a'^2}{16}(2+3e^2)(1-3\cos^2 i)$$
$$+ \omega_z H' - \frac{3}{128}\frac{\mu^6 R_m^4}{L'^3 G'^7} J_4 [3-30\cos^2 i+35\cos^4 i](2+3e^2)$$
$$- \frac{n_E^2 a'^2}{16} \sin^2 I_E [(2+3e^2)\tfrac{3}{2}(3\cos^2 i-1)] \qquad (16)$$

The coupling terms in Eq. (14) have been dropped because of their relatively small size. All of the terms independent of g and h in F_3^* are included in F_2^*. F_3^* is

$$F_3^* = -\frac{3}{8}\frac{\mu^5 R_m^3}{L'^3 G'^5} J_3 e' \sin i\,(1-5\cos^2 i)\sin g'$$
$$+ \frac{30}{128}\frac{\mu^6 R_m^4}{L'^3 G'^7} J_4[1-8\cos^2 i+7\cos^4 i]e^2 \cos 2g'$$
$$+ \frac{15}{16} n_E^2 a'^2 e'^2 \sin^2 i' \cos 2g' - \omega_y G' \sin i' \cos h'$$
$$+ \frac{n_E^2 a'^2}{16}\Bigg[\sin I_E \Big\{(6+9e'^2)\sin 2i' \cos h'$$
$$-30e'^2 \sin i' [\cos i' \cos h' \cos 2g' -$$
$$\sin h' \sin 2g']\Big\}$$
$$\sin^2 I_E \Big\{(3+\tfrac{9}{2}e'^2)\sin^2 i' \cos 2h'$$
$$-15e'^2 \cos i' \sin 2h' \sin 2g'$$
$$+ \tfrac{15}{2}e'^2(1+\cos^2 i')\cos 2h' \cos 2g'$$
$$- \tfrac{45}{2}e'^2 \sin^2 i' \cos 2g'\Big\}\Bigg] \qquad (17)$$

It should be pointed out that for the lower orbits the various angles appearing in Eq. (17) i.e., g', h', $h' \pm g'$, $h' \pm 2g'$, are all driven by the dominant F_2^* term. One must beware however of the various instances where any of the angular rates become small. This occurs near the eleven (slightly altitude dependent) critical inclinations[6] at i' = 46.6, 56.1, 63.4, 69.0, 73.2, 90.0, 106.8, 111.0, 116.6, 123.9, and 133.6 degrees. When the inclination is near one of these critical values, the resulting behavior of the coordinates can exhibit very long period variations and the von Zeipel method, now to be followed, fails. A method valid in these special situations will be outlined in a later section. For a high orbiter having a moderate inclination (above 40 deg.) F_2^* in Eq. (16) will contain some g' and h' dependent terms that are now included in F_3^*. In this case the von Zeipel procedure fails. This situation is discussed by Kozai[1] and Vagners[9].

IV. THE LONG PERIOD TERMS

As before one looks for a transformation from variables (L',G',H',g',h') to new variables (L",G",H",g",h") such that the new Hamiltonian F^{**} is a function of (L",G",H") only. Consider the stationary generating function

$$S^* = L''\ell' + G''g' + H''h' + S_1^*(L'',G'',H'',g'',h'') + \ldots \qquad (18)$$

The relation between the coordinates are

$$G' = G'' + \frac{\partial S_1^*}{\partial g'} + \ldots \qquad g'' = g' + \frac{\partial S_1^*}{\partial G''} + \ldots \qquad (19)$$

with similar relations between the other variables.

Choose S_1^* to be of the form

$$S_1^* = \alpha_1 \sin 2g' + \alpha_2 \cos g' + \alpha_3 \sin h' + \alpha_4 \sin 2h' + \alpha_5 \sin(h'+2g') + \alpha_6 \sin(h'-2g') + \alpha_6 \sin(2h'+2g') + \alpha_7 \sin(2h'-2g') \qquad (21)$$

where

$$\alpha_1 = \left[\frac{-30}{128} \frac{\mu^6 R_m^4}{L''^3 G''^7} J_4 \left\{ 1-8\cos^2 i''+7\cos^4 i'' \right\} e''^2 - \frac{15}{16} n_E^2 a''^2 e''^2 \sin^2 i'' \left(1 - \frac{3}{2}\sin^2 I_E\right) \right] \Big/ 2\frac{\partial F_2^*}{\partial G''}$$

$$\alpha_2 = \left[-\frac{3}{8} \frac{\mu^5 R_m^3}{L''^3 G''^5} J_3 \, e'' \sin i'' (1-5\cos^2 i'') \right] \Big/ \frac{\partial F_2^*}{\partial G''}$$

$$\alpha_3 = \left[+ \omega_z G'' \sin i'' - \frac{n_E^2 a''^2}{16} \sin I_E (6+9e''^2) \sin 2i'' \right] \Big/ \frac{\partial F_2^*}{\partial H''}$$

$$\alpha_4 = \left[-\frac{n_E^2 a''^2}{64} \sin^2 I_E (6+9e''^2) \sin^2 i'' \right] \Big/ \frac{\partial F_2^*}{\partial H''}$$

$$\alpha_5 = \left[-\frac{15}{16} n_E^2 a''^2 e''^2 \sin i_E \sin i'' (1+\cos i'') \right] \Big/ \left[2\frac{\partial F_2^*}{\partial G''} + \frac{\partial F_2^*}{\partial H''} \right]$$

$$\alpha_6 = \left[-\frac{15}{16} n_E^2 a''^2 e''^2 \sin I_E \sin i'' (1-\cos i'') \right] \Big/ \left[2\frac{\partial F_2}{\partial G''} - \frac{\partial F_2^*}{\partial H''} \right]$$

$$\alpha_7 = \left[-\frac{15}{64} n_E^2 a''^2 e''^2 \sin^2 I_E (1+\cos i'')^2 \right] \Big/ \left[2\frac{\partial F_2^*}{\partial G''} + 2\frac{\partial F_2^*}{\partial H''} \right]$$

$$\alpha_8 = \left[+ \frac{15}{64} n_E^2 a''^2 e''^2 \sin^2 I_E (1-\cos i'')^2 \right] \Big/ \left[2\frac{\partial F_2^*}{\partial G''} - 2\frac{\partial F_2^*}{\partial H''} \right]$$

The Hamiltonian F** is now independent of g" and h" and is

$$F^{**}(L'',G'',H'') = \frac{\mu^2}{2L''^2} + \frac{1}{4}\frac{\mu^4 R_m^2}{L''^3 G''^3} J_{20}(1-3\cos^2 i'') + \omega_z H''$$
$$- \frac{3}{128}\frac{\mu^6 R_m^4}{L''^3 G''^7} J_4(2+3e''^2)\left[3-30\cos^2 i''+35\cos^4 i''\right]$$
$$+ \frac{n_E^2 a''^2}{16}(2+3e''^2)(3\cos^2 i''-1)(1-\frac{3}{2}\sin^2 I_E) \qquad (22)$$

As S_1 and S_1^* are known, Eqs. (11) and (19) are utilized to determine the behavior of the elememts (L,G,H,ℓ,g,h). The coefficients of the trigonometric terms in Eq. (21) contain six critical divisors. These correspond to the eleven (slightly altitude dependent) critical inclinations mentioned previously. It appears from an inspection of Eq. (21) that near a critical i" the amplitude of the coordinate variations can become nearly infinite. Actually this is not the case as will be shown in the following example.

Suppose the inclination is near 90 degrees. A near polar orbit is chosen as the very long period behavior resulting from the laws of Cassini is best demonstrated. The slowly varying Hamiltonian is (i.e., the relatively fast variable, g', has been averaged out)

$$F^{**} = \frac{\mu^2}{2L''^2} + \frac{1}{4}\frac{\mu^4 R_m^2}{L''^3 G''^3} J_{20}(1-3\cos^2 i'') + \omega_z H'' - \omega_y \sin i''\cos h''$$
$$+ \frac{n_E^2 a''^2}{16}(2+3e''^2)\Big[(3\cos^2 i''-1) + 3\sin I_E \sin 2i''\cos h''$$
$$+ \frac{3}{2}\sin^2 I_E \sin^2 i''\cos 2h''\Big] \qquad (23)$$

where all of the elements are slowly varying. The secular terms involving J_4 and $\sin^2 I_E$ have been omitted in Eq. (23) due to their small size. Since the Hamiltonian does not depend on ℓ'' or g", L" and G" are constant and the equations of motion have been reduced to that of one degree of freedom. They are expressed as

$$\dot{h}'' = -\frac{\partial F^{**}}{\partial H''} = \left[\frac{3}{2} J_{20}\left(\frac{R_m}{p''}\right)^2 - \frac{3}{8}\left(\frac{n_E}{n}\right)^2 \frac{(2+3e^2)}{\sqrt{1-e^2}}\right] n \frac{H''}{G''} - \omega_z \qquad (24)$$

$$-G''\sin i''\frac{di''}{dt} = \dot{H}'' = +\frac{\partial F^{**}}{\partial h''} = \omega_y G''\sin i''\sin h''$$

$$- \frac{3}{16} n_E^2 a''^2 (2+3e''^2)\left[\sin I_E \sin 2i''\sin h'' + \sin^2 I_E \sin^2 i''\sin 2h''\right] \qquad (25)$$

Note that $\omega_z = -1.07\times10^{-8}$ and $J_{20} = -2.0\times10^{-4}$. The phase plane (H",h") contours of constant F** in Eq. (23) are shown in Fig. (3). For very low orbits (Fig. 3a) the stable equilibrium points occur at h" = 0 degrees and $i'' = \cos^{-1}\left(\frac{H''}{G''}\right) = 88$ degrees. Recall that the nodal position, h", is measured from the point where the plane of the lunar equator, the ecliptic, and the terrestrial path meet in a common line. It appears that the orbital plane of a low orbiter can become trapped in this same configuration. Or it can exhibit very slow stable oscillations about this position, the period of which is about twenty years for a low satellite.

The interplay between the inclination and nodal position demonstrates remarkably different behavior for higher orbiters. For a semi-major axis of 1.5 R_m and an eccentricity of 0.0, the behavior of (H",h") or equivalently (i",Ω'') is shown in Fig. (3b). In this case the inclination can be trapped near 83 degrees but h" appears to be stable between 0 and 90 degrees. The behavior for a still higher orbiter is shown in Fig. (3c).

For this case the stable equilibrium solutions occur at Ω'' slightly over 90 degrees (and slightly less than 270 degrees) and $i'' = 74$ degrees. This is near the 73.4 degree critical inclination (when $\dot{h}'' - \dot{g}'' = 0$) mentioned previously. However in this particular case a closer examinination shows that this critical i'' occurs at about 66 degrees. This particular orbit may eventually impact the surface. (cf. Kozai)

V. CONCLUSIONS

The long period behavior of a lunar orbiter is determined for a certain class of orbits. The method of successive approximations is employed in treating the circulating orbits. In this case the angles g'', h'', $g''+h''$, etc., were assumed to move at nearly uniform rates. The librating orbits (when one of the angles does not move through an angle of 360 degrees) are treated by the use of a phase plane analysis. Treated, as an example of the latter, are near polar orbits in which the indirect effect of the Sun (described by the laws of Cassini) is important. Stable altitude dependent positions of the orbital plane are found.

VI. ACKNOWLEDGEMENTS

The author wishes to thank his advisor, Professor John V. Breakwell, for his helpful suggestions and guidance.

This research was supported by the National Aeronautics and Space Administration under Contract No. NsG 133-61.

REFERENCES

1. Kozai, Y., "Motion of a Lunar Orbiter," Journal of the Astronomical Society of Japan, 15, 1963.

2. Giacaglia, G.E.O., "The Motion of a Satellite of the Moon", Goddard Space Flight Center Report X-547-65-218, Greenbelt, Maryland, June 1965.

3. Osterwinter, C., "The Motion of a Lunar Satellite," Ph.D. Dissertation, Yale University 1965.

4. Melbourne, W.G., et al, "Constants and Related Information for Astrodynamic Calculations," J.P.L. Technical Report 32-1306, July 1968.

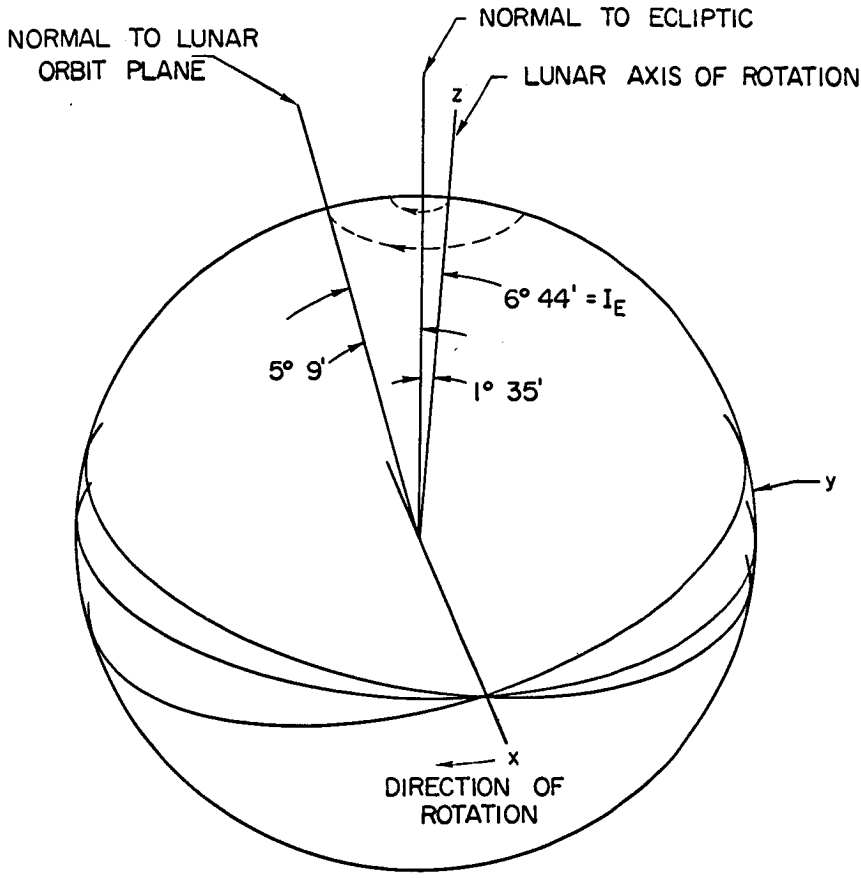
5. Akim, E.L., "Determination of the Gravity Field of the Moon by the Motion of the AMS LUNA 10," N.A.S.A. No. ST-CM-LPS-10532.

6. Breakwell, J.V. and D. Hensley., "An investigation of High Eccentricity Orbits about Mars," First Compilation of Papers on Trajectory Analysis and Guidance Theory, N.A.S.A. SP-141, 1967.

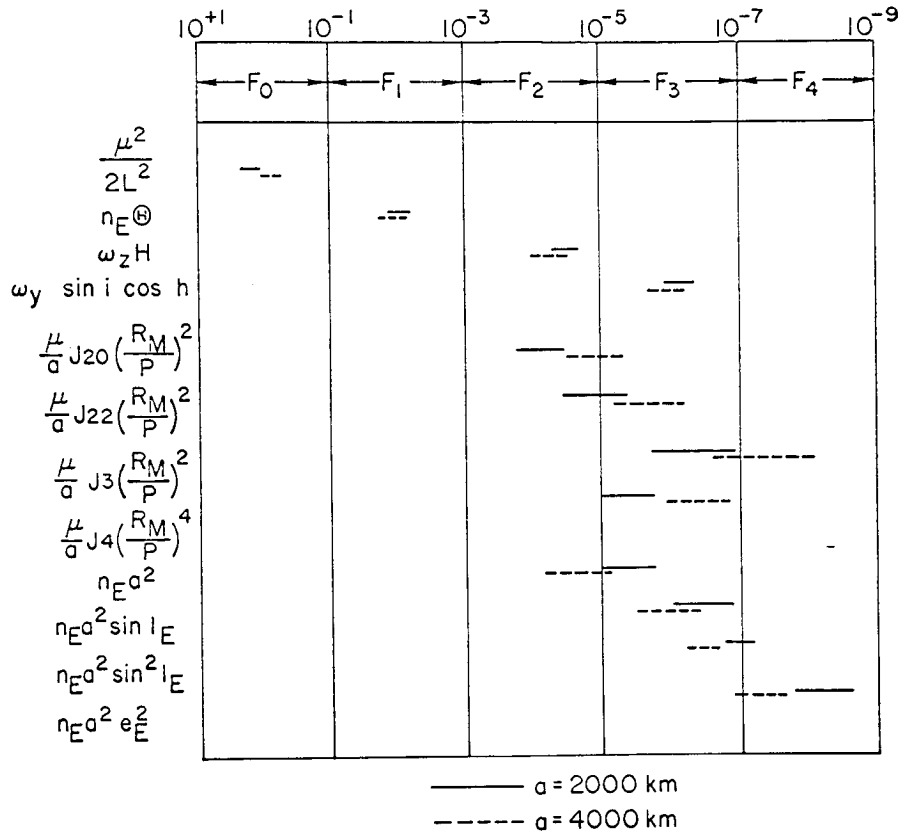
7. Brouwer, D., "Solution of the Problem of Artifical Satellites Without Drag," Astronomical Journal, 64, 1959.

8. Garfinkel, B., "On the Motion of a Satellite in the Vicinity of the Critical Inclination," Astronomical Journal, 65, 1960.

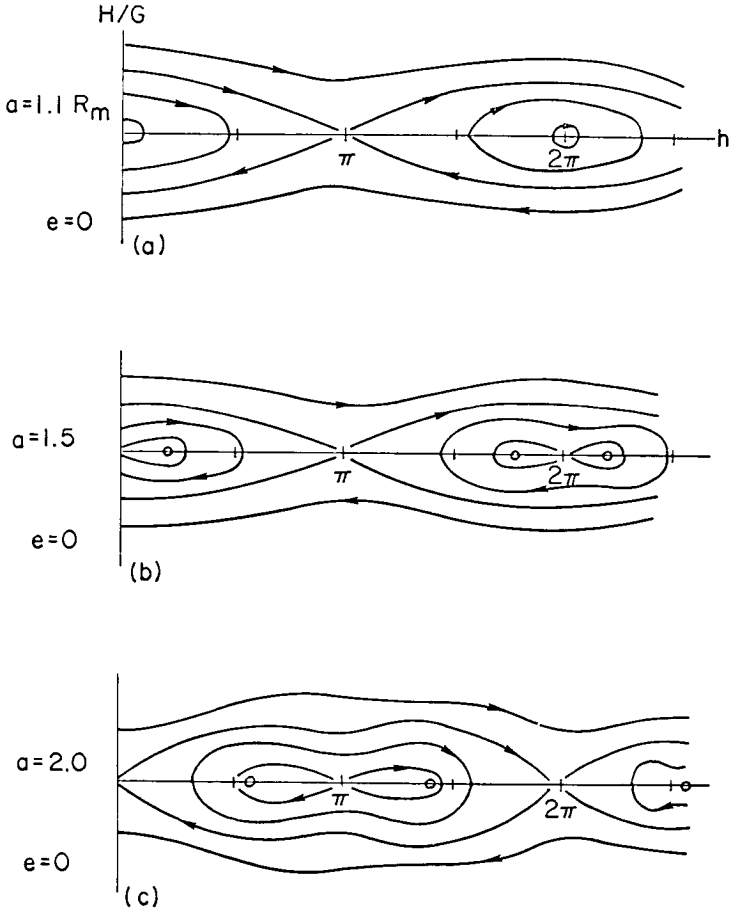
9. Vagners, J., "Some Resonant and Non-Resonant Perturbations of Earth and Lunar Orbiters," Ph.D. Dissertation, Stanford University, August 1967.



SYSTEM OF REFERENCE



THE RELATIVE IMPORTANCE OF THE VARIOUS EARTH MOON PERTURBATIONS



PHASE SPACE CONTOURS

LECTURES ON NONLINEAR RESONANCE

By W. T. Kyner
University of Southern California

LECTURES ON NONLINEAR RESONANCE

W. T. KYNER

In the early 1940's, interest in the theory of nonlinear differential equations developed rapidly in the United States. Friedrichs, Hurewitz, Levinson, Stoker at Brown, Lefschetz, Bellman at Princeton, and Minorsky at the David Taylor Model Basin were among those most responsible. In particular, Lefschetz recognized the importance of the Soviet contributions during the preceding decade and helped make much of this work accessible to the American technical public. In 1942, he prepared a translation of excerpts from monographs of Krylov and Bogoljubov [5] whose averaging techniques are closely related to the general perturbation theories of celestial mechanics. It is interesting to note, however, that just when Krylov and Bogoljubov were starting their research in nonlinear mechanics, an elderly American, E. W. Brown, Gibbs professor of mathematics at Yale University, explained and essentially justified the important concept of resonance as a basically nonlinear phenomenon. His lectures, "Elements of the Theory of Resonance Illustrated by the Motion of a Pendulum," were given at the Rice Institute in April 1931 and were later published as a Rice Institute pamphlet [2]. They are particularly relevant to this year's Yale Summer Institute because of the importance of resonance phenomena in geodetic satellite theory.

In my lectures on resonance I shall follow Brown's exposition of the basic concepts, but I shall use the Krylov-Bogoljubov method of averaging in the mathematical analysis. The main application of this theory will be to satellite problems.

LECTURES ON NONLINEAR RESONANCE

1. Pendulum problems

As we all know, a stretched wire has certain modes of vibration which seem independent of the strength of the energy source. But we tend to forget that the "natural frequencies" of these modes are a mathematical fiction since they are only present "when the vibrations have infinitely small amplitudes, which amounts to saying that the wire is not vibrating at all. More properly, a natural frequency should be defined as the lower limit of the frequency of that particular mode of vibration. It is necessary to insist on this change of frequency with change of amplitude because the existence of the phenomena of resonance depends on the existence of this change" (p. 2 of [5]). Furthermore, a detailed analysis of the "locking in" effect which is observed when two piano wires are tuned to the same frequency depends in an essential way on the change of frequency with amplitude. This is discussed in detail by Brown and more concisely by Cesari (p. 151 of [3]). I shall omit such a discussion here and go directly to the pendulum problems which are physically less interesting, but more relevant to satellite theory.

We first consider an ideal pendulum of length b with an oscillating support (see figure 1).

Let Y be the horizontal distance of the support point S from a fixed point O and x the angle which the pendulum makes with the vertical. The support point S is constrained to move in the horizontal direction. The equation of motion of the pendulum is

$$\frac{d^2Y}{dt^2}\cos x + b\frac{d^2x}{dt^2} = -g\sin x, \tag{1.1}$$

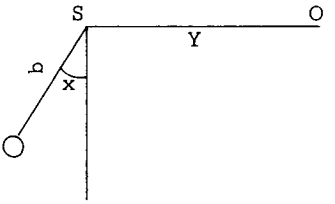


FIG. 1

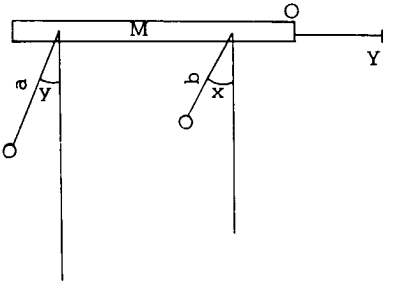


FIG. 2

(we sum the forces along the line perpendicular to the pendulum).

We now assume that S oscillates with a motion given by $Y(t)$ where

$$\frac{d^2Y}{dt^2} = -\varepsilon g f(\alpha t), \quad \varepsilon \text{ small}, \quad f(z + 2\pi) = f(z), \quad \text{all } z.$$

Then

$$\frac{d^2x}{dt^2} + \omega^2 \sin x = \varepsilon f(\psi) \cos x, \qquad \omega^2 = g/b, \qquad \psi = \alpha t. \tag{1.2}$$

The second model problem is that of two pendulums, each of mass m, but with different lengths, attached to a bar of mass M which is constrained to move in the horizontal direction (see figure 2). We assume that the total horizontal momentum is zero, i.e.,

$$\frac{d}{dt}[MY + m(Y + b \sin x) + m(Y + a \sin y)] = 0. \tag{1.3}$$

The equations of motion are

$$\begin{aligned} \frac{d^2Y}{dt^2} \cos x + b \frac{d^2x}{dt^2} &= -g \sin x, \\ \frac{d^2Y}{dt^2} \cos y + a \frac{d^2y}{dt^2} &= -g \sin y. \end{aligned} \tag{1.4}$$

Using (1.3) to eliminate d^2Y/dt^2, we obtain

$$\frac{d^2x}{dt^2} + \omega^2 \sin x = \varepsilon \cos x \frac{d^2}{dt^2}(\sin x + \frac{a}{b} \sin y),$$

(1.5)

$$\frac{d^2y}{dt^2} + \alpha^2 \sin y = \varepsilon \cos y \frac{d^2}{dt^2}(\sin y + \frac{b}{a} \sin x),$$

where $\omega^2 = g/b$, $\alpha^2 = g/a$, $\varepsilon = m(M + 2m)^{-1}$. We assume that ε is small. Equations (1.5) are awkward to work with since the second derivatives of x and y appear in both equations. We therefore rewrite the equations as

$$\frac{d^2x}{dt^2} + \omega^2 \sin x = -\varepsilon \cos x [1 - \varepsilon (\cos^2 x + \cos^2 y)]^{-1}$$
$$\{\omega^2 (\cos x \sin x + \cos y \sin y) + \sin x (\frac{dx}{dt})^2$$
$$+ \frac{a}{b} \sin y (\frac{dy}{dt})^2\},$$

(1.6)

$$\frac{d^2y}{dt^2} + \alpha^2 \sin y = -\varepsilon \cos y [1 - \varepsilon (\cos^2 x + \cos^2 y)]^{-1}$$
$$\{\alpha^2 (\cos x \sin x + \cos y \sin y) + \frac{b}{a} \sin x (\frac{dx}{dt})^2$$
$$+ \sin y (\frac{dy}{dt})^2\}.$$

Each equation of (1.6) can be interpreted as a perturbed pendulum equation. We therefore can use the same mathematical procedures on equations (1.2) and (1.6). The first, and rather difficult, step is to introduce new coordinates so that the differential equations

will be in the normal form for the method of averaging. In order to motivate the coordinate change and to display the simplest features of resonance, we shall now study the linear differential equation obtained from (1.2) by making the small angle approximation, i.e., $\sin x \doteq x$, $\cos x \doteq 1$. We have

$$(1.7) \qquad \frac{d^2x}{dt^2} + \omega^2 x = \varepsilon f(\psi), \qquad \psi = \alpha t.$$

The solution to (1.7) can be written

$$(1.8) \qquad x(t) = r_0 \cos \theta(t) + \frac{\varepsilon}{\omega} \int_0^t \sin \omega(t - u) f(\alpha u)\, du,$$

where $\theta(t) = \omega t + \theta_0$, r_0 and θ_0 are constants determined by initial conditions.

If α, the frequency of the forcing function, is an integral multiple of ω, the frequency of the linearized pendulum equation, then unbounded solutions of (1.7) are possible. In other words, if

$$(1.9) \qquad \alpha k = \omega, \qquad k \text{ an integer},$$

then the condition of linear resonance has been satisfied. It is obviously the same for all forcing functions of period $2\pi/\alpha$, but the existence of unbounded solutions depends on the presence of $\sin \omega t$ or $\cos \omega t$ in the Fourier series expansion of a particular forcing function. The concept of linear resonance is of limited physical significance since the small angle approximation is destroyed.

Before leaving the linear approximation, let us consider the

homogeneous ($\varepsilon = 0$) problem with the aid of the corresponding phase and potential planes (see figure 3). In the phase plane we plot the level curves of the energy integral,

$$E(x,\dot{x}) = \frac{1}{2}\dot{x}^2 + \frac{1}{2}\omega^2 x^2 = h, \quad \text{a constant},$$

and in the potential plane, the two curves

$$c = h, \qquad c = \frac{1}{2}\omega^2 x^2.$$

Each level curve is characterized by its energy and therefore by its amplitude. We now introduce r, the amplitude (note that $h = \omega^2 r^2/2$), and θ, a normalized angle, as dependent variables, i.e., we set

$$x = r\cos\theta, \quad \dot{x} = -\omega r \sin\theta, \quad \theta = \omega t + \theta_0. \tag{1.10}$$

The inhomogeneous equation (1.7) is equivalent to

$$\begin{aligned} \frac{dr}{dt} &= \frac{\varepsilon}{\omega} f(\psi)\sin\theta, \\ \frac{d\theta}{dt} &= \omega + \frac{\varepsilon}{\omega r} f(\psi)\cos\theta, \\ \frac{d\psi}{dt} &= \alpha. \end{aligned} \tag{1.11}$$

Equations (1.11) are in the normal form for the method of averaging. Note that if ε is nonzero, they are nonlinear.

In order to reduce the nonlinear pendulum equation (1.2) to normal form, we seek a coordinate transformation (compare with (1.10))

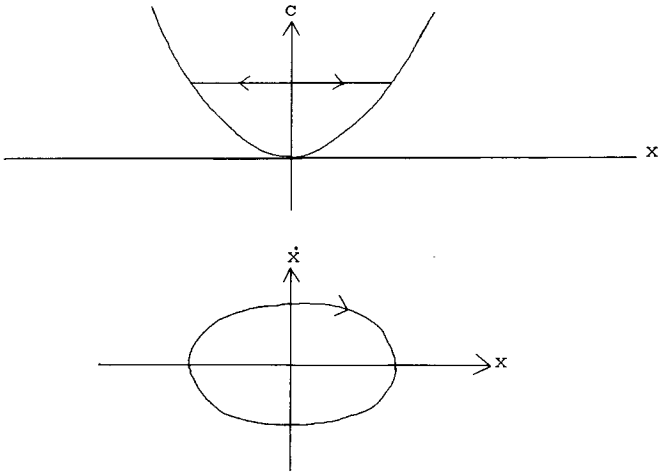


FIG. 3

(1.12) $x = F(r,\theta), \quad \dot{x} = G(r,\theta),$

with F and G having period 2π in θ, such that the unperturbed ($\varepsilon = 0$) equations have the form

$$\frac{dr}{dt} = 0,$$

(1.13) $$\frac{d\theta}{dt} = z(r),$$

$$\frac{d\psi}{dt} = \alpha.$$

It will be shown later that $z(r) = \omega^2(1 - r^2/16 + \ldots)$.

The construction of the transformation (1.12) is somewhat complicated, but it can be motivated with the aid of the phase and potential planes. We again plot (see figure 4) the level curves of the energy integral,

$$E(x,\dot{x}) = \frac{1}{2}\dot{x}^2 + \omega^2(1 - \cos x) = h, \text{ a constant},$$

in the phase plane, and the curves

$$c = h, \quad c = \omega^2(1 - \cos x),$$

in the potential plane.

We see that h less than $2\omega^2$ implies that the motion is periodic and that the level curves are characterized by the amplitude. The transformation (1.12) is therefore possible. Since we need to study several nonlinear differential equations in the satellite problems,

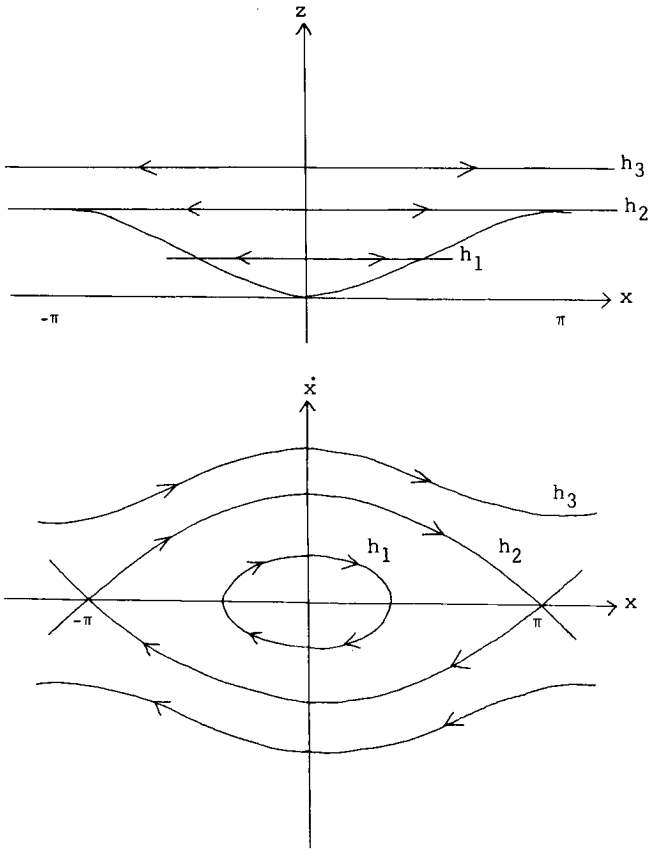


FIG. 4

we shall give a general construction which will then be applied to the pendulum problems. It is similar to the one used by Brown and more recently by Morgunov [7].

Consider

$$(1.14)\qquad \frac{d^2x}{dt^2} + q(x) = 0, \qquad q(-x) = -q(x), \qquad q'(0) > 0.$$

If r is positive and not too large, then

$$(1.15)\qquad \frac{1}{2}\dot{x}^2 + Q(x) = Q(r), \quad \text{where } Q(x) = \int_0^x q(x')\,dx',$$

is the equation of a closed integral curve in the phase plane. It is generated by a periodic solution of (1.14).

From

$$(1.16)\qquad \dot{x} = (2[Q(r) - Q(x)])^{1/2} = A(r,x),$$

we have

$$(1.17)\qquad \theta = z(r) \int^x A^{-1}(r,x')\,dx' = B(r,x),$$

where the frequency $z(r)$ is given by

$$(1.18)\qquad 2\pi\, z^{-1}(r) = 4 \int_0^r A^{-1}(r,x')\,dx' .$$

Equations (1.16-1.18) implicitly define the required transformation (1.12). Clearly,

$$(1.19) \qquad \frac{d\theta}{dt} = z(r), \qquad \frac{dr}{dt} = 0.$$

The perturbed equation,

$$(1.20) \qquad \frac{d^2x}{dt^2} + q(x) = \varepsilon f(\alpha t),$$

will be transformed into a system of first order equations,

$$\frac{d\theta}{dt} = z(r) + \varepsilon \Theta(r,\theta,\psi),$$

$$(1.21) \qquad \frac{dr}{dt} = \varepsilon R(r,\theta,\psi),$$

$$\frac{d\psi}{dt} = \alpha.$$

To do this we write

$$\frac{d\dot{x}}{dt} = \frac{\partial A}{\partial r}\frac{dr}{dt} + \frac{\partial A}{\partial x}\frac{dx}{dt},$$

(1.22)

$$\frac{d\theta}{dt} = \frac{\partial B}{\partial r}\frac{dr}{dt} + \frac{\partial B}{\partial x}\frac{dx}{dt}.$$

But

$$\frac{dx}{dt} = A(r,x),$$

$$\frac{d^2x}{dt^2} = -q(x) + \varepsilon f(\alpha t),$$

and

$$\frac{\partial A}{\partial r} = q(r)\, A^{-1}(r,x), \quad \frac{\partial A}{\partial x} = -q(x)\, A^{-1}(r,x),$$

$$\frac{\partial B}{\partial x} = z(r)\, A^{-1}(r,x).$$

Hence,

$$\frac{d\theta}{dt} = z(r) + \varepsilon\, f(\psi)\, A(r,x)\, z^{-1}(r)\, \frac{\partial B}{\partial r}(r,x),$$

$$\frac{dr}{dt} = \varepsilon\, f(\psi)\, A(r,x)\, q^{-1}(r), \tag{1.23}$$

$$\frac{d\psi}{dt} = \alpha,$$

where $\theta = B(r,x)$ determines x as a function of θ and r.

Equation (1.23) can be simplified if we have an explicit formula for $x = F(r,\theta)$, the inverse to $\theta = B(r,x)$. For from

$$\dot{x} = A(r,x) = z(r)\, \frac{\partial F}{\partial \theta}(r,\theta),$$

$$A(r,x)\, \frac{\partial B}{\partial r}(r,x) = -A(r,x)\, \frac{\partial B}{\partial x}(r,x) = -z(r)\, \frac{\partial F}{\partial r}(r,\theta),$$

we obtain

$$\frac{dr}{dt} = \varepsilon\, f(\psi)\, z(r)\, q^{-1}(r)\, \frac{\partial F}{\partial \theta}(r,\theta),$$

$$\frac{d\theta}{dt} = z(r)\, [1 - \varepsilon\, f(\psi)\, q^{-1}(r)\, \frac{\partial F}{\partial r}(r,\theta)], \tag{1.24}$$

$$\frac{d\psi}{dt} = \alpha.$$

Note that $F(r,\theta)$ and $z(r)$ can frequently be represented by an infinite series, e.g., by using Lindstedt's method (p. 116 of Cesari [3]).

We have finally transformed our differential equations into the normal form for the method of averaging, our next topic.

2. The method of averaging.

For convenience, new notation is employed in this section. We consider a system of ordinary differential equations,

$$\frac{dx}{dt} = \varepsilon X(x,y), \qquad x = (x_1,\ldots,x_M),$$

(2.1)

$$\frac{dy}{dt} = z(x) + \varepsilon Y(x,y), \qquad y = (y_1,\ldots,y_N),$$

with initial conditions $x(0) = a$, $y(0) = b$. The vector valued functions, $X(x,y)$, $Y(x,y)$, are assumed to be smooth and to have period 2π in each y_n. The x_m are called slow variables, the y_n fast variables, since if $\varepsilon = 0$,

$$x_m = a_m, \qquad 1 \leq m \leq M,$$

(2.2)

$$y_n = z_n(a)\, t + b_n, \qquad 1 \leq n \leq N.$$

Our goal is to construct a transformation,

$$x = u + \varepsilon P(u,v),$$

(2.3)

$$y = v + \varepsilon Q(u,v),$$

with P and Q having period 2π in each v_n, so that the equations (2.1) become

$$\frac{du}{dt} = \varepsilon U(u) + \varepsilon^2 W_1(u,v,\varepsilon),$$

(2.4)

$$\frac{dv}{dt} = z(u) + \varepsilon V(u) + \varepsilon^2 W_2(u,v,\varepsilon).$$

In other words, the fast variables have been eliminated (to first order) from the differential equations. As we shall now show, this elimination is an averaging procedure; in fact, we can take

$$U(u) = (2\pi)^{-N} \int_0^{2\pi} \cdots \int_0^{2\pi} X(u,y)\, dy_1 \cdots dy_N,$$

(2.5)

$$V(u) = (2\pi)^{-N} \int_0^{2\pi} \cdots \int_0^{2\pi} Y(u,y)\, dy_1 \cdots dy_N .$$

Approximate solutions to (2.1) can be constructed by solving the first order averaged equations,

$$\frac{du}{dt} = \varepsilon\, U(u), \qquad u(0) = a',$$

$$\frac{dv}{dt} = z(u) + \varepsilon\, V(u), \qquad v(0) = b', \tag{2.6}$$

$$a = a' + \varepsilon\, P(a',b'), \quad b = b' + \varepsilon\, Q(a',b'),$$

and substituting the solution into (2.3).

If we differentiate (2.3), then from (2.1) and (2.4), we have

$$\varepsilon\, X(u + \varepsilon\, P,\ v + \varepsilon\, Q) = (I_M + \varepsilon\, \frac{\partial P}{\partial u})(\varepsilon\, U + \varepsilon^2 W_1)$$

$$+ \varepsilon\, \frac{\partial P}{\partial v}\,(z(u) + \varepsilon\, V + \varepsilon^2 W_2),$$

$$z(u + \varepsilon\, P) + \varepsilon\, Y(u + \varepsilon\, P,\ v + \varepsilon\, Q) = (I_N + \varepsilon\, \frac{\partial Q}{\partial v})(z(u)$$

$$+ \varepsilon\, V + \varepsilon^2 W_2)$$

$$+ \varepsilon\, \frac{\partial Q}{\partial u}\,(\varepsilon\, U + \varepsilon^2 W_1).$$

Expanding in powers of ε, we have

$$\varepsilon\, X(u,v) = \varepsilon[U(u,v) + \frac{\partial P}{\partial v}\,(u,v)\ z(u)] + \varepsilon^2[**],$$

$$z(u) + \varepsilon[\frac{\partial z}{\partial u}\,(u)\ P(u,v) + Y(u,v)] = z(u) + \varepsilon[V(u)$$

$$+ \frac{\partial Q}{\partial v}\,(u,v)\ z(u)] + \varepsilon^2[**],$$

where $[**]$ denotes a smooth function of u, v, ε whose explicit formula is not needed here. Clearly, we must require that

$$X(u,v) - U(u) = \frac{\partial P}{\partial v}(u,v)\, z(u),$$

(2.7)

$$\frac{\partial z}{\partial u}(u)\, P(u,v) + Y(u,v) - V(u) = \frac{\partial Q}{\partial v}(u,v)\, z(u)\ .$$

If the first equation is to have a periodic solution, the left side must have zero mean. From this requirement, we have the first equation of (2.5). If, in addition, we know that $P(u,v)$ has zero mean, then in order to solve for $Q(u,v)$, we must have the second equation of (2.5).

Let us briefly consider vector equations of the type

$$F(u,v) = \frac{\partial S}{\partial v}(u,v)\, z(u), \tag{2.8}$$

where the given function $F(u,v)$ has period 2π in each v_n. Since we seek a periodic solution, we expand both S and F in a Fourier series,

$$S = \sum_{\underline{j}} S_{\underline{j}}(u) \exp i[\underline{j},v], \qquad \underline{j} = (j_1,\ldots,j_N),$$

$$F = \sum_{\underline{j}} F_{\underline{j}}(u) \exp i[\underline{j},v], \tag{2.9}$$

where

$$[\underline{j},v] = \sum_1^N j_n v_n\ .$$

Substituting (2.9) into (2.8) and equating the coefficients of $\exp i[\underline{j},v]$, we obtain an infinite set of equations,

$$F_{\underline{0}}(u) = 0,$$

(2.10)

$$F_{\underline{j}}(u) = i[\underline{j},z(u)]\, S_{\underline{j}}(u) .$$

If for all u in the domain of interest, and for all integer vectors $\underline{j}$, we have the nonresonance condition,

(2.11) $\quad [\underline{j},z(u)] \neq 0,$

then we can solve for the $S_{\underline{j}}(u)$, and therefore for $S(u,v)$. We obtain a formal series,

(2.12) $\quad S(u,v) = \sum_{\underline{j}\neq\underline{0}} - i[\underline{j},z(u)]^{-1}\, F_{\underline{j}}(u) \exp i[\underline{j},v] .$

The denominators $[\underline{j},z(u)]$ can become small as $\underline{j}$ becomes large, thereby preventing the convergence of the series (2.12). We avoid this difficult problem (the classical small divisors problem) by assuming that $F(u,v)$, and therefore, $S(u,v)$ are trigonometric polynomials. Note that to the particular solution of zero mean (2.12) we can add an arbitrary solution of the homogeneous equation,

(2.13) $\quad 0 = \frac{\partial S}{\partial v}(u,v)\, z(u) .$

Returning to (2.4), we see that if the nonresonance condition (2.11) is satisfied, and if U and V are chosen by (2.5), then the transformation (2.3) can be constructed.

It is not always convenient to require that $P(u,v)$ and

$Q(u,v)$ have zero mean. For example, if the system (2.1) is Hamiltonian, then the solution of the homogeneous equation (2.13) can be selected so that the averaged equations (2.6) are Hamiltonian. John Morrison [8] has developed a generalized method of averaging in which the additive arbitrary functions play an essential role. Much of this section is based on his work.

Before discussing the resonance problems (our main topic), we shall investigate in what sense the solutions to (2.6) determine approximate solution to the original equations (2.1). For simplicity, we shall consider scalar equations, i.e., $N = M = 1$.

Let $u(t)$ and $v(t)$ be solutions of the exact equations (2.4), and $u^*(t)$, $v^*(t)$ be solutions of the approximate equations (2.6) with

$$u(0) = u^*(0) = a', \quad v(0) = v^*(0) = b'.$$

Let

$$r = u - u^*, \quad s = v - v^*, \tag{2.14}$$

$$x^* = u^* + \varepsilon P(u^*,v^*), \quad y^* = v^* + \varepsilon Q(u^*,v^*).$$

Then

$$\frac{dr}{dt} = \varepsilon [U(u^* + r) - U(u^*)] + \varepsilon^2 W_1,$$

$$\frac{ds}{dt} = z(u^* + r) - z(u^*) + \varepsilon [V(u^* + r) - V(u^*)] + \varepsilon^2 W_2. \tag{2.15}$$

We assume that in the domain of interest the given functions and their derivatives can be bounded by the same constant C. Then from (2.15), we obtain the inequality

$$|r(t)| \leq \varepsilon \int_0^t C|r(t')|\, dt' + \varepsilon^2 Ct ,$$

(2.16)

$$|s(t)| \leq (1 + \varepsilon) \int_0^t C|r(t')|\, dt' + \varepsilon^2 Ct .$$

By the generalized Gronwall inequality (p. 11 of Sanone and Conti [9]), we have

$$|r(t)| \leq \varepsilon [\exp(\varepsilon Ct) - 1] \leq \varepsilon^2 Ct \exp(\varepsilon Ct),$$

(2.17)

$$|s(t)| \leq (1 + \varepsilon)[\exp(\varepsilon Ct) - 1] + \varepsilon^2 Ct - \varepsilon C(1 + \varepsilon)\, t^2/2$$

$$\leq \varepsilon Ct\, [2 \exp(\varepsilon Ct) + \varepsilon - t] .$$

Therefore from (2.3) and (2.14), we have the error estimates,

$$|x(t) - x^*(t)| \leq (1 + \varepsilon C)\, |r(t)| + \varepsilon C\, |s(t)| ,$$

(2.18)

$$|y(t) - y^*(t)| \leq (1 + \varepsilon C)\, |s(t)| + \varepsilon C\, |r(t)| .$$

On an interval $0 \leq t \leq T/\varepsilon$, T fixed, the error in the slow variable satisfies the inequality,

(2.19) $\quad |x(t) - x^*(t)| \leq \varepsilon^2 t\, C^*,$

while the error in the fast variable satisfies the weaker inequality,

$$|y(t) - y^*(t)| \leq \varepsilon\, t\, C^* , \tag{2.20}$$

where the constant C^* depends on the bounds C and T.

In general, the estimates (2.19) and (2.20) are the best possible. Approximations which are meaningful on an infinite time interval can be constructed only under the most exceptional circumstances.

We now have developed the mathematical machinery for studying nonlinear resonance.

DEFINITION. If there exists an x_0 and $\underline{k} \neq 0$ such that

$$[\underline{k}, z(x_0)] = 0, \tag{2.21}$$

then the condition for resonance motion has been satisfied. The degree of the resonance is the number of linearly independent integer vectors $\underline{k}$ which satisfy (2.21).

A resonant problem can be reduced to a nonresonant problem by suitably reducing the number of fast variables. Let L denote the module of integer vectors $\underline{k}$ such that $[\underline{k}, z(x_0)] = 0$, and let

$$\underline{k}_1, \ldots, \underline{k}_\nu, \qquad 1 \leq \nu \leq N$$

be a basis of L. We can construct $N - \nu$ linearly independent vectors

$$\underline{k}_{\nu+1}, \ldots, \underline{k}_N,$$

perpendicular to L. Hence

(2.22) $\Delta = \det |\underline{k}_1,\ldots,k_N| \neq 0.$

Set

(2.23) $q_n = \Delta^{-1} [\underline{k}_n,y]$ or $q = Ky$,

where the $N \times N$ matrix K has as its rows the vectors $\underline{k}_n/\Delta$.

By construction,

(2.24) $[\underline{j},Kz(x_0)] = 0$

implies that the last $N - \nu$ components of $\underline{j}$ are zero.

The next step is similar to that used in boundary layer studies. We set

(2.25) $x = x_0 + \varepsilon^{1/2} p$.

By virtue of (2.23) and (2.25), the original system of differential equations (2.1) is equivalent to

$$\begin{aligned} \frac{dp}{dt} &= \varepsilon^{1/2} X(x_0 + \varepsilon^{1/2} p,K^{-1}q) , \\ \frac{dq}{dt} &= Kz(x_0 + \varepsilon^{1/2} p) + \varepsilon KY(x_0 + \varepsilon^{1/2}p,K^{-1}q) . \end{aligned} \tag{2.26}$$

Hence

$$\begin{aligned} \frac{dp}{dt} &= \varepsilon^{1/2} X(x_0,K^{-1}q) + \varepsilon X_1(p,q,\varepsilon^{1/2}) , \\ \frac{dq}{dt} &= Kz(x_0) + \varepsilon^{1/2} D(x_0) p + \varepsilon Y_1(p,q,\varepsilon) , \end{aligned} \tag{2.27}$$

where $D(x_0) = K\partial z(x_0)/\partial x$, etc.

Since the components of K^{-1} are integers, the system (2.27) has period 2π in each q_n. Furthermore $q_1,\ldots,q_\nu$ as well as $p_1,\ldots p_M$ are slow variables, while $q_{\nu+1},\ldots,q_N$ are fast variables.

Two features of the transformation of (2.1) to (2.27) should be emphasized. The equations (2.27) are valid in a neighborhood of x_0; quoting from Brown (p. 8 of [2]) "resonance is not a single special case of motion but is a group of cases extending over a finite range of values of the constants." The second feature is a drop in the order of the approximation; $\varepsilon^{1/2}$ rather than ε is the perturbation parameter.

We make one last change in notation. Let

$$\mu = \varepsilon^{1/2} , \qquad D(x_0) = \begin{pmatrix} \Lambda \\ \Phi \end{pmatrix},$$

$$\lambda = (q_1,\ldots,q_\nu) , \qquad X(x_0, K_q^{-1}) = R(\lambda,\phi), \text{ etc.}$$

(2.28)

$$\phi = (q_{\nu+1},\ldots,q_N) ,$$

$$\omega = ([\underline{k}_{\nu+1},z(x_0)],\ldots,[\underline{k}_N,z(x_0)]) ,$$

Then

$$\frac{dp}{dt} = \mu R(\lambda,\phi) + \mu^2 R_1(p,\lambda,\phi,\mu) ,$$

(2.29)
$$\frac{d\lambda}{dt} = \mu \Lambda p + \mu^2 \Lambda_1(p,\lambda,\phi,\mu) ,$$

$$\frac{d\phi}{dt} = \omega + \mu \Phi p + \mu^2 \Phi_1(p,\lambda,\phi,\mu) .$$

Clearly, the first order (in μ) averaged equations are (we forego another change of notation)

$$\frac{dp}{dt} = \mu R_0(\lambda), \quad R_0(\lambda) = (2\pi)^{-(N-\nu)} \int_0^{2\pi} \cdots \int_0^{2\pi} R(\lambda,\phi)\, d\phi_1 \cdots d\phi_{N-\nu}$$

$$\frac{d\lambda}{dt} = \mu \Lambda p, \tag{2.30}$$

$$\frac{d\phi}{dt} = \omega + \mu \Phi p .$$

Note that in both (2.29) and (2.30) Λ and Φ are constant matrices. Furthermore, we have the nonresonance condition $[\underline{j},\omega] = 0$ implies $\underline{j} = \underline{0}$.

Let us consider the first two equations of (2.30).

$$\frac{dp}{dt} = \mu R_0(\lambda) ,$$

$$\frac{d\lambda}{dt} = \mu \Lambda p . \tag{2.31}$$

If $R_0(\lambda)$ vanishes at $\lambda = \lambda_0$, then $p = 0$, $\lambda = \lambda_0$ is an equilibrium state of the system (2.31). The stability of the equilibrium state can be studied with the aid of the linearized equations,

$$\frac{dp}{dt} = \mu \frac{\partial}{\partial \lambda} R_0(\lambda_0)\, (\lambda - \lambda_0) ,$$

$$\frac{d}{dt}(\lambda - \lambda_0) = \mu \Lambda p . \tag{2.32}$$

The phenomenon of libration occurs if the equilibrium state is the center of a family of periodic solutions of (2.31). Just as with the linearized pendulum equations (1.7), the condition of resonance does not by itself insure that the motion has any special properties. For example, if the frequency of the unperturbed problem does not change with the amplitude, then Λ is zero, and libration is impossible. Hence Brown's claim, "the existence of the phenomena of resonance depends on the existence of this change."

Before returning to the pendulum examples, one final observation should be made. The first order system (2.32) is equivalent to a second order system,

$$\frac{d^2\lambda}{dt^2} = \mu^2 \Lambda R_0(\lambda) . \tag{2.33}$$

If there exists a scalar function $Z(\lambda)$ such that

$$\Lambda R_0(\lambda) = - \operatorname{grad} Z(\lambda) ,$$

then

$$\frac{1}{2}\left[\frac{d\lambda}{dt}, \frac{d}{dt}\lambda\right] + \mu^2 Z(\lambda) = \text{const.} \tag{2.34}$$

is an integral of (2.33).

3. Analysis of the pendulum problems

As our first example, let us take the linearized pendulum equations (see (1.7) and (1.11)),

$$\frac{dr}{dt} = -\frac{\varepsilon}{\omega} f(\psi) \sin\theta ,$$

$$\frac{d\theta}{dt} = \omega + \frac{\varepsilon}{\omega r} f(\psi) \cos\theta , \tag{3.1}$$

$$\frac{d\psi}{dt} = \alpha .$$

Note that we must have r nonzero.

If

$$\omega = k\alpha, \quad k \text{ an integer}, \tag{3.2}$$

the condition for linear resonance is satisfied.

Following the procedure of the preceding section, we set

$$\mu = \varepsilon^{1/2} ,$$

$$\lambda = (\theta - k\psi)/(1 + k^2) ,$$

$$\phi = (k\theta + \psi)/(1 + k^2) , \tag{3.3}$$

$$r = r_0 + \mu p, \quad r_0 > 0.$$

Then

$$\theta = \lambda + k\phi ,$$

$$\psi = -k\lambda + \phi ,$$

and

$$\frac{dp}{dt} = -\frac{\mu}{\omega} f(\phi - k\lambda) \sin(\lambda + k\phi) ,$$

$$\frac{d\lambda}{dt} = \frac{\mu^2 f(\phi - k\lambda) \cos(\lambda + k\phi)}{\omega(r_0 + \mu p)(1 + k^2)} , \tag{3.4}$$

$$\frac{d\phi}{dt} = +\frac{\mu^2 k\, f(\phi - k\lambda) \cos(\lambda + k\phi)}{\omega(r_0 + \mu p)(1 + k^2)} .$$

The first order (in μ) averaged equations are

$$\frac{dp}{dt} = -\frac{\mu}{\omega} \frac{1}{2\pi} \int_0^{2\pi} f(\phi - k\lambda) \sin(\lambda + k\phi)\, d\phi ,$$

$$\frac{d\lambda}{dt} = 0 , \tag{3.5}$$

$$\frac{d\phi}{dt} = \alpha .$$

For simplicity, let us take $f(\psi) = \cos k\psi$. From

$$\cos(k\phi - k^2\lambda) \sin(\lambda + k\phi) = \frac{1}{2} \sin(1 + k^2)\lambda + \frac{1}{2} \sin[2k\phi + (1 - k^2)\lambda] ,$$

we have that the equations for the slow variables p and λ are

$$\frac{dp}{dt} = -\frac{\mu}{2\omega} \sin(1 + k^2)\lambda ,$$

$$\frac{d\lambda}{dt} = 0 . \tag{3.6}$$

Clearly, there are no periodic solutions of (3.6). It is important to note that the (unstable) equilibrium states are not of physical significance.

The second example is the nonlinear pendulum equations (see (1.2) and (1.24)). In normal form, we have

$$\frac{dr}{dt} = \varepsilon f(\psi)\, z(r)\, \frac{\partial F}{\partial \theta}(r,\theta)[\omega^2 \sin r]^{-1} ,$$

$$\frac{d\theta}{dt} = z(r) - \varepsilon\, z(r)\, f(\psi)\, \frac{\partial F}{\partial r}(r,\theta)[\omega^2 \sin r]^{-1} , \tag{3.7}$$

$$\frac{d\psi}{dt} = \alpha ,$$

where

$$z(r) = \omega\left(1 - \frac{r^2}{16}\right) + O(r^4) ,$$

$$x = F(r,\theta) = r \cos\theta + O(r^2) . \tag{3.8}$$

It is essential that $dz/dr \neq 0$.

We shall not derive (3.8) in detail, but only remark that if we set

$$\sin x/2 = \sin r/2 \sin \zeta ,$$

then (see (1.18))

$$z(r) = \omega\left[\frac{1}{\pi}\int_0^{\pi} (1 - \sin^2 r/2 \sin^2 \zeta)^{-1/2}\, d\zeta\right]^{-1}$$

We can now easily derive the first equation of (3.8) by expanding the integrand in powers on r.

Unfortunately, the transformation from x, $\dot{x}$ to r, θ is singular at $r = 0$. For simplicity, we want r small, but if the equations (3.7) are to be meaningful, we must have r nonzero. We therefore assume that

$$0 < \varepsilon^{\gamma} \leq r^2 \leq \varepsilon^{\gamma'} . \tag{3.9}$$

The exponents of the bounds on r^2 will be chosen shortly.

We now replace (3.7) by the simplified equations,

$$\begin{aligned} \frac{dr}{dt} &= -\frac{\varepsilon}{\omega} f(\psi) \sin\theta + O(\varepsilon r^2) , \\ \frac{d\theta}{dt} &= \omega\left(1 - \frac{r^2}{16}\right) + \frac{\varepsilon}{\omega r} f(\psi) \cos\theta + O(\varepsilon r^2) , \\ \frac{d\psi}{dt} &= \alpha . \end{aligned} \tag{3.10}$$

The difference, and it is essential, between (3.1) and (3.10) is that the frequence of θ depends on r.

The condition for <u>nonlinear resonance</u> is that

$$\omega(1 - r_0^2/16) = k\alpha , \quad k \text{ an integer}, \quad r_0 \neq 0 . \tag{3.11}$$

Assuming (3.11), we make the substitution (3.3) and obtain

$$\frac{dp}{dt} = \frac{\mu}{\omega} f(\phi - k\lambda) \sin(\lambda + k\phi) + O(\mu r^2) ,$$

$$(3.12) \qquad \frac{d\lambda}{dt} = -\frac{\mu\omega r_0 p}{8(1+k^2)} - \frac{\mu^2\omega p^2}{16(1+k^2)} + \frac{\mu^2 f(\phi-k\lambda)\sin(\lambda+k\phi)}{\omega(r_0+\mu p)(1+k^2)} + O(\mu^2 r^2),$$

$$\frac{d\phi}{dt} = \alpha + \frac{k}{(1+k^2)}\left[-\frac{\mu\omega}{8} r_0 p - \frac{\mu^2\omega}{16} p^2 + \frac{\mu^2 f(\phi-k\lambda)\sin(\lambda+k\phi)}{\omega(r_0+\mu p)}\right.$$

$$\left. + O(\mu^2 r^2)\right] .$$

In the $d\lambda/dt$ equation we want the first term to be dominant even though r_0 is small, i.e., we want $\mu r_0 \gg \mu^2/r_0$. Therefore, we require that

$$(3.13) \qquad 0 < \mu^{1-\gamma} \leq r^2 \leq \mu^{1-\gamma/2} , \qquad 0 < \gamma < 1/2 .$$

The upper bound permits us to drop the $O(\mu r^2)$ and $O(\mu^2 r^2)$ terms in (3.12). Clearly, from (3.13), we obtain (3.9)

We again set $f(\psi) = \cos k\psi$. The first order averaged equations are

$$\frac{dp}{dt} = -\frac{\mu}{2\omega} \sin(1+k^2)\lambda ,$$

$$(3.14) \qquad \frac{d\lambda}{dt} = -\frac{\mu\omega r_0 p}{8(1+k^2)} ,$$

$$\frac{d\phi}{dt} = \alpha - \frac{\mu\omega k r_0 p}{8(1+k^2)}$$

Let $\zeta = (1 + k^2)\lambda + \pi$, $\tau = \mu t$. Then the differential equations for the slow variables λ and p are equivalent to a homogeneous pendulum equation,

$$\frac{d^2\zeta}{dt^2} + \Omega^2 \sin \zeta = 0, \qquad \Omega^2 = r_0/8 . \tag{3.15}$$

Note that the independent variable is the "slow time," $\tau = \mu t$, and that the frequency Ω depends on the amplitude r_0.

The example of the pendulum with oscillating support can be discussed in more detail (see section III of Brown [2]), but we have displayed its most important properties. To summarize: because the frequency is amplitude dependent, libration can occur at resonance. The equations describing this libration are equivalent to the homogeneous pendulum equations, but are valid only over a finite time interval (of the order of $\varepsilon^{-1/2}$).

The two pendulum problem gives similar results. Here we set

$$\begin{aligned} x &= r_1 \cos \theta_1 + O(r_1^2), \qquad z_1(r_1) = \omega(1 - r_1^2/16), \\ y &= r_2 \cos \theta_2 + O(r_2^2), \qquad z_2(r_2) = \alpha(1 - r_2^2/16), \\ f_1 &= -\omega^2(r_1 \cos \theta_1 + r_2 \cos \theta_2), \\ f_2 &= -\alpha^2(r_1 \cos \theta_1 + r_2 \cos \theta_2). \end{aligned} \tag{3.16}$$

Then equations (1.6) are approximated by

$$\frac{d}{dt} r_1 = - \varepsilon\omega (r_1 \cos\theta_1 + r_2 \cos\theta_2) \sin\theta_1 ,$$

$$\frac{d}{dt} r_2 = - \varepsilon\alpha (r_1 \cos\theta_1 + r_2 \cos\theta_2) \sin\theta_2 ,$$

(3.17)

$$\frac{d}{dt} \theta_1 = \omega (1 - r_1^2/16) + \frac{\varepsilon\omega}{r_1} (r_1 \cos\theta_1 + r_2 \cos\theta_2) \cos\theta_1 ,$$

$$\frac{d}{dt} \theta_2 = \alpha (1 - r_2^2/16) + \frac{\varepsilon\alpha}{r_2} (r_1 \cos\theta_1 + r_2 \cos\theta_2) \cos\theta_2 .$$

The condition for nonlinear resonance is

(3.18) $$k_1\omega (1 - r_{10}^2/16) = k_2\alpha (1 - r_{20}^2/16) .$$

We shall take $k_1 = k_2 = 1$. The analysis of other resonances is left to the reader.

Then with

$$\mu = \varepsilon^{1/2} , \quad \lambda = \frac{1}{2} (\theta_1 - \theta_2) , \quad \phi = \frac{1}{2} (\theta_1 + \theta_2) ,$$

(3.19)

$$r_1 = r_{10} + p_1 , \quad r_2 = r_{20} + \mu p_2 ,$$

we have (retaining the dominant terms)

$$\frac{d}{dt} p_1 = - \frac{\mu\omega}{2}[r_{10} \sin 2(\lambda+\phi) + r_{20} \sin 2\phi + r_{20} \sin 2\lambda] ,$$

$$\frac{d}{dt} p_2 = - \frac{\mu\alpha}{2}[r_{20} \sin 2(\lambda+\phi) + r_{10} \sin 2\phi - r_{10} \sin 2\lambda] ,$$

(3.20)

$$\frac{d\lambda}{dt} = \frac{\mu}{16} [\alpha r_{20} p_2 - \omega r_{10} p_1] ,$$

$$\frac{d\phi}{dt} = \alpha(1 - r_{20}^2/16) - \frac{\mu}{16} (\omega r_{10} p_1 + \alpha r_{20} p_2) .$$

The averaged equations are

$$\frac{d}{dt} p_1 = -\frac{\mu\omega}{2} r_{20} \sin 2\lambda ,$$

$$\frac{d}{dt} p_2 = \frac{\mu\alpha}{2} r_{10} \sin 2\lambda ,$$

(3.21)

$$\frac{d\lambda}{dt} = \frac{\mu}{16} [\alpha r_{20} p_2 - \omega r_{10} p_1] ,$$

$$\frac{d\phi}{dt} = \alpha (1 - r_{20}^2/16 - \frac{\mu}{16} (\omega r_{10} p_1 + \alpha r_{20} p_2) .$$

Once again with $\zeta = 2\lambda + \pi$, $\tau = \mu t$, we have the homogeneous pendulum equation describing the resonance phenomena,

$$\frac{d^2\zeta}{dt^2} + \Omega^2 \sin \zeta = 0, \qquad \Omega^2 = (\alpha^2 + \omega^2) r_{10} r_{20}/16 . \tag{3.22}$$

The libration around $\zeta = 0$ corresponds to an exact solution of the original equations (1-6) in the special case of equal lengths, i.e., $\alpha = \omega$. For then, if $x \equiv -y$, the two pendulums oscillate out of phase with exactly the same frequency. The support is motionless. If the lengths are almost equal, this "locking in" can be approximated if the initial displacements are chosen so that the resonance condition (3.18) is satisfied. The unstable equilibrium point, $\zeta = \pi$, corresponds to initial conditions $x \cong y \neq 0$. The support must then oscillate (preservation of linear momentum)--the subsequent motion of the pendulums is eratic. Finally, we note that the simplicity of equations (3.15) and (3.22) is due to the small amplitude assumption.

4. Motion of synchronous satellites.

In this section, we shall study a typical satellite problem in order to establish the accuracy and the time interval of validity of the pendulum model which is derived by the method of averaging. The problem is the determination of the effects of asymmetries in the Earth's gravitational field on the motion of a synchronous satellite, i.e., one whose mean motion is approximately equal to the rotation rate of the Earth. Because of the near equality of the two frequencies, the mean motion and the rotation rate, we have an example of nonlinear resonance where the effects of the longitude dependent asymmetries are amplified. This resonance has been carefully studied by L. Blitzer [1], B. Morando [6] and others. W. Kaula's textbook "Theory of Satellite Geodesy" [4] contains a clear exposition of the phenomenon.

For simplicity, we shall ignore those terms in the potential which have little effect on synchronous satellites and write the potential as

$$
\begin{aligned}
V &= -\frac{\mu}{r} - V_{20} - V_{22}, \\
V_{20} &= -\frac{J_2}{r^3} P_2(\sin\varphi), \\
V_{22} &= +\mu \frac{J_{22}}{r^3} P_{22}(\sin\varphi) \cos 2(\lambda - \lambda_{22}),
\end{aligned}
\tag{4.1}
$$

where

μ = the gravitational constant,

r = the radial distance from the center of mass measured in Earth radii,

φ = the latitude,

$$\lambda = \text{the longitude},$$

$$J_2 \approx 10^{-3}$$

$$J_{22} \approx 10^{-6}$$

$$P_2(\sin\varphi) = \frac{1}{2}(3\sin^2\varphi - 1),$$

$$P_{22}(\sin\varphi) = 3(1 - \sin^2\varphi)$$

Then, according to the standard theory, (Chapter 3 of [4]) if the mean secular rate (short periodic effects are suppressed) of λ_A, the astromical latitude, is close to zero then the secular behavior of λ_A is governed by the pendulum equation, i.e.,

$$\ddot{\lambda}_A = A(a, e, i) \sin 2(\lambda_A - \lambda_{22}) \tag{4.2}$$

where

$$\lambda_A = \omega + M + \Omega - \theta$$

$$\theta = \gamma t + \epsilon_o, \text{ Greenwich sideral time,}$$

$$A(a, e, i) = \frac{\mu}{a^5} J_{22} \frac{9}{2}(1 + \cos i)^2 \left(1 - \frac{5}{2}e^2 + \frac{13}{16}e^4 + \ldots\right).$$

Here the variables $(a, e, i, \omega, \Omega, M)$ are to be interpreted as <u>averaged</u> or <u>mean</u> elements. It should be noted that the dominate asymmetry, V_{20}, does not influence the secular behavior of λ_A. Furthermore, if

$$\dot{\lambda}_A = \dot{\omega} + \dot{\Omega} + \dot{M} - \dot{\theta} \approx 0 \tag{4.3}$$

and if J_2 and J_{22} are zero, then $\dot{M} - \dot{\theta} \approx 0$, i.e., we have near equality of the mean motion of the satellite and the rotation rate of the Earth. As usual

a = the semi-major axis,

e = the eccentricity,

i = the inclination,

ω = the argument of perigee,

Ω = the longitude of the ascending node,

M = the mean anomaly.

We shall start with the equations of motion and derive averaged equations corresponding to the pendulum equations (4.2) for a special class of orbits, namely, nearly circular orbits in the equatorial plane. This restriction is convenient because we want to consider the longitude dependent term V_{22} as the perturbation and J_{22} as the perturbation parameter. In other words, the unperturbed potential is

$$V = \frac{-\mu}{r} + \frac{\mu}{r^3} J_2 P_2(\sin\varphi).$$

For equatorial orbits, $\sin\varphi \equiv 0$, and the unperturbed problem is an integrable central force problem. If we make the additional restriction that the unperturbed orbit is geometrically circular, then the algebraic details will not obscure our purpose, the application of the method of averaging to typical nonlinear resonance problems.

With the center of mass of the Earth as the origin of an inertial coordinate system, we set $x = r\cos w$, $y = r\sin w$. Then

$$T = \frac{1}{2}[\dot{r}^2 + r^2\dot{w}^2], \text{ the kinetic energy,}$$

$$V = \frac{-\mu}{r} - \frac{\mu}{2r^3} J_2 - \frac{\mu}{r^3} 3J_{22}\cos 2\lambda, \text{ the potential energy,} \tag{4.4}$$

$$L = T - V, \text{ the Lagrangian,}$$

where (compare with (4.1)) we have set $\sin\varphi \equiv 0$, $\lambda_{22} = 0$, and $\lambda = w - \theta$.

The equations of motion are

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{r}} - \frac{\partial L}{\partial r} = 0$$

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{w}} - \frac{\partial L}{\partial w} = 0$$

or

$$\frac{d^2r}{dt^2} - r\left(\frac{dw}{dt}\right)^2 + \frac{\mu}{r^2} = -\mu\frac{3}{2}\frac{J_2}{r^4} - \frac{\mu 9 J_{22}}{r^4}\cos 2\lambda$$

$$\frac{d}{dt}\left(r^2\frac{dw}{dt}\right) = \frac{-\mu 6 J_{22}}{r^3}\sin 2\lambda \tag{4.5}$$

We now take w as the independent variable and λ, c, u, du/dw as dependent variables, where

$$\lambda = w - \theta,\quad c = r^2\frac{dw}{dt},\quad u = \frac{1}{r},\quad \frac{du}{dw} = -\frac{\dot{r}}{c}\ . \tag{4.6}$$

We verify that

$$\frac{d^2u}{dw^2} + u = \frac{\mu}{c^2} + \frac{3}{2}\frac{\mu}{c^2}J_2u^2 + \frac{\mu}{c^2}9J_{22}u^2\cos 2\lambda - \frac{1}{c}\frac{du}{dw}\frac{dc}{dw},$$

$$\frac{dc}{dw} = \frac{-\mu 6 J_{22}}{c}u\sin 2\lambda, \tag{4.7}$$

$$\frac{d\lambda}{dw} = 1 - \frac{\gamma}{cu^2}\ .$$

The equations of motion have equilibrium solutions which correspond to circular orbits with 24-hour periods. They are determined by the transcendental equations obtained by setting the

derivatives equal to zero in (4.7),

$$u = \frac{\mu}{c^2}\left[1 + \frac{3}{2} J_2 u^2 + 9 J_{22} u^2 \cos 2\lambda\right],$$

$$(4.8) \qquad 0 = -\frac{\mu}{c} 6u J_{22} \sin 2\lambda,$$

$$0 = 1 - \frac{\gamma}{cu^2}.$$

Hence,

$$\lambda = 0, \pm \frac{\pi}{2}, \pm \pi,$$

$$c = \frac{\gamma}{u^2},$$

and u is the positive solution of

$$u = \frac{\mu u^4}{\gamma^2}\left[1 + \frac{3}{2} J_2 u^2 + 9 J_{22} u^2\right].$$

If we were only interested in studying solutions near the equilibrium solutions, then we would introduce normal coordinates relative to each of the solutions of $\sin 2\lambda = 0$. However, if λ is to be unrestricted, it is more convenient to set $J_{22} = 0$ in (4.8) and define our unperturbed orbit by the equations,

$$(4.9) \qquad u_o = \frac{\mu}{c_o^2}\left[1 + \frac{3}{2} J_2 u_o^2\right], \quad \gamma = c_o u_o^2.$$

It should be noted that this orbit has nonzero instantaneous eccentricity, but is geometrically circular.

We now introduce variables p_1, p_2, p_3 which correspond to deviations from the resonant amplitudes of section 2. Let

$$\frac{\mu}{c^2} = \frac{\mu}{c_o^2} + \sqrt{J_{22}}\, p_3 ,$$

$$u = u_o + \sqrt{J_{22}}[p_3 + p_1 \cos w + p_2 \sin w], \tag{4.10}$$

$$\frac{du}{dw} = 0 + \sqrt{J_{22}}[0 - p_1 \sin w + p_2 \cos w].$$

Then, using (4.9), we find that

$$\frac{dp_3}{dw} = \sqrt{J_{22}}\, B(p_1, p_2, p_3, \lambda, w; J_2, J_{22})$$

$$\frac{dp_1}{dw} \cos w + \frac{dp_2}{dw} \sin w = \sqrt{J_{22}}\, C(p_1, p_2, p_3, \lambda, w; J_2, J_{22}), \tag{4.11}$$

$$-\frac{dp_1}{dw} \sin w + \frac{dp_2}{dw} \cos w = \sqrt{J_{22}}\, D(p_1, p_2, p_3, \lambda, w; J_2, J_{22}),$$

or

$$\frac{dp_1}{dw} = \sqrt{J_{22}}\, \{C \cos w - D \sin w\} ,$$

$$\frac{dp_2}{dw} = \sqrt{J_{22}}\, \{C \sin w + D \cos w\} \tag{4.12}$$

$$\frac{dp_3}{dw} = \sqrt{J_{22}}\, B ,$$

where

$$B = -C = 12\frac{\mu^2}{c^4} u \sin 2\lambda ,$$

(4.13)

$$J_{22}D = \frac{d^2u}{dw^2} + u - \frac{\mu}{c^2} - u_o + \frac{\mu}{c_o^2}$$

$$= \frac{3}{2} J_2 \left(\frac{\mu}{c^2} u^2 - \frac{\mu}{c_o^2} u_o^2\right)$$

$$+ 9 J_{22} \frac{\mu}{c^2} u^2 \cos 2\lambda$$

$$- \frac{1}{c} \frac{du}{dw} \frac{dc}{dw} .$$

From (4.7) and (4.10), we have

$$\frac{\mu}{c^2} u^2 - \frac{\mu}{c_o^2} u_o^2 = \sqrt{J_{22}} \{p_3 u_o^2 + 2u_o \frac{\mu}{c_o^2} (p_3 + \delta)$$

$$+ \sqrt{J_{22}} \frac{\mu}{c_o^2} (p_3 + \delta)^2 + J_{22} p_3 (p_3 + \delta)^2\} ,$$

$$- \frac{1}{c} \frac{du}{dw} \frac{dc}{dw} = \sqrt{J_{22}} \{6 J_{22} \left(\frac{\mu}{c_o^2} + \sqrt{J_{22}}\, p_3\right)\left(u_o + \sqrt{J_{22}} (p_3 + \delta)\right)$$

$$\cdot (-p_1 \sin w + p_2 \cos w) \sin 2\lambda\} ,$$

with $\delta = p_1 \cos w + p_2 \sin w$.

Therefore,

$$\sqrt{J_{22}}\, D = J_2 \{(u_o^2 + 2u_o \frac{\mu}{c_o^2}) p_3 + 2u_o \frac{\mu}{c_o^2}(p_1 \cos w + p_2 \sin w)$$

$$+ (\sqrt{J_{22}} \frac{\mu}{c_o^2} + J_{22} p_3)(p_3 + p_1 \cos w + p_2 \sin w)^2\}$$

$$+ 9 \sqrt{J_{22}}(\frac{\mu}{c_o^2} + \sqrt{J_{22}}\, p_3)(u_o + \sqrt{J_{22}}\, (p_3 + p_1 \cos w$$

(4.14) $$+ p_2 \sin w)^2 \cos 2\lambda + 6\, J_{22}(\frac{\mu}{c_o^2} + \sqrt{J_{22}} p_3)(u_o + \sqrt{J_{22}} \cdot$$

$$(p_3 + p_1 \cos w + p_2 \sin w))\ \ (-p_1 \sin w + p_2 \cos w) \sin 2\lambda$$

$$= J_2 \{(u_o^2 + 2\, u_o \frac{\mu}{c_o^2})\, p_3 + 2\, u_o \frac{\mu}{c_o^2} (p_1 \cos w + p_2 \sin w)\}$$

$$+ 9 \sqrt{J_{22}} \frac{\mu}{c_o^2}\, u_o \cos 2\lambda + 0\, (J_2 \sqrt{J_{22}} + J_{22}).$$

After simplification, we have

$$\frac{dp_1}{dw} = -\sqrt{J_{22}} \{9 \frac{\mu}{c_o^2}\, u_o \cos 2\lambda \sin w$$

$$+ 12 \frac{\mu^2}{c_o^4}\, u_o \sin 2\lambda \cos w\}$$

$$-J_2 \{(u_o^2 + 2\, u_o \frac{\mu}{c_o^2})\, p_3 \sin w$$

$$+ 2\, u_o \frac{\mu}{c_o^2} (p_1 \cos w + p_2 \sin w) \sin w\} + 0\, (J_2 \sqrt{J_{22}} + J_{22}),$$

(4.15)

$$\frac{dp_2}{dw} = \sqrt{J_{22}} \{9 \frac{\mu}{c_o^2}\, u_o \cos 2\lambda \cos w$$

(cont'd)

$$- 12 \frac{\mu^2}{c_o^2} u_o \sin 2\lambda \sin w\}$$

$$+ J_2 \{(u_o^2 + 2 u_o \frac{\mu}{c_o^2}) p_3 \cos w$$

$$+ 2 u_o \frac{\mu}{c_o^2} (p_1 \cos w + p_2 \sin w) \cos w\}$$

$$+ 0 (J_2 \sqrt{J_{22}} + J_{22}),$$

$$\frac{dp_3}{dw} = \sqrt{J_{22}} \{ 12 \frac{\mu^2}{c_o^4} u_o \sin 2\lambda \} + 0 (J_{22}).$$

Furthermore, from (4.7), (4.9), and (4.10), we have

$$\frac{d\lambda}{dt} = 1 - \frac{\gamma}{cu^2} = 1 - \frac{c_o u_o^2}{cu^2}$$

$$= 1 - \left(1 + \sqrt{J_{22}} \frac{c_o^2}{\mu} p_3\right)^{\frac{1}{3}} \left(1 + \frac{\sqrt{J_{22}}}{u_o} (p_3 + \delta)\right)^{-2} \tag{4.16}$$

$$= \sqrt{J_{22}} \{(\frac{2}{u_o} - \frac{c_o^2}{2\mu}) p_3 + \frac{2}{u_o} (p_1 \cos w + p_2 \sin w)\}$$

$$+ 0 (J_{22})$$

Since $J_2 \approx \sqrt{J_{22}}$, and

$$\frac{dw}{dt} = c u^2 = \gamma + 0 (\sqrt{J_{22}}), \tag{4.17}$$

w is a fast variable, and p_1, p_2, p_3, λ are slow variables. We now average with respect to w and use (4.17) to obtain the first order (in $\sqrt{J_{22}}$) averaged equations,

$$\frac{dp_1}{dt} = - J_2 \{ \gamma u_o \frac{\mu}{c_o^2} p_2 \},$$

$$\frac{dp_2}{dt} = J_2 \{ \gamma u_o \frac{\mu}{c_o^2} p_1 \},$$

(4.18)

$$\frac{dp_3}{dt} = \sqrt{J_{22}} \{ \gamma 12 \frac{\mu}{c_o^2} u_o \sin 2\lambda \},$$

$$\frac{d\lambda}{dt} = \sqrt{J_{22}} \{ \gamma (\frac{2}{u_o} - \frac{c_o^2}{2\mu}) p_3 \}.$$

The qualitative effect of V_{22}, the longitude dependent term in the potential, is now easy to describe. From (4.18), we have

$$p_1 = s \cos \eta (w - w_o), \quad p_2 = s \sin \eta (w - w_o),$$

(4.19)

$$p_3 = \sqrt{J_{22}} \gamma 12 \frac{\mu^2}{4 c_o} u_o \int^w \sin 2\lambda (w') \, dw',$$

where $\eta = J_2 \gamma u_o \mu / c_o^2$, s, w_o are integration constants, and λ is a solution of the pendulum equation,

$$\frac{d^2\lambda}{dt^2} = G(\gamma, J_2, J_{22}) \sin 2\lambda,$$

where

$$G = J_{22} 12 \gamma^2 \frac{\mu^2}{4 c_o} (2 - \frac{c_o^2 u_o}{\mu}),$$

$$c_o = \frac{\gamma}{u_o^2},$$

and u_o is the positive function of γ and J_2 determined by (4.9).

From (4.10), we have

$$\frac{1}{r} = \frac{1}{r_o} + \sqrt{J_{22}}\,[p_3 + p_1 \cos w + p_2 \sin w].$$

Therefore, the radial distance will oscillate about its mean value. The behavior of λ, the longitude of the satellite, is more interesting since it is strongly influenced by initial conditions. If λ is near one of the unstable equilibrium values (0 or π), or if p_3 is large enough, the satellite will slowly drift around the Earth. On the other hand, if λ is near one of the stable equilibrium values ($+\pi$ or $-\pi$) and if p_3 is small enough, then the satellite will librate about the equilibrium value. The libriation period can be shown to be proportional to $1/\sqrt{J_{22}}$; the constant of proportionality is dependent on the initial conditions.

Finally, we note that the pendulum equations (4.2) and (4.20) are compatible since

$$\frac{c_o^2}{\mu} = a_o(1 - e_o^2),\ \frac{1}{u_o} = a_o(1 - e_o),\ \mu = \gamma^2 a_o^3$$

and from (4.9) it follows that $e_o = 0(J_2)$.

In conclusion, we note that since the pendulum equation (4.20) was derived by the method of averaging, the error estimates of section 2 are applicable and we can now assert that the pendulum model is valid over a time interval proportional to $1/\sqrt{J_{22}}$, e.g., over a libration period. Statements of this type must be accompanied by the phrase "if J_{22} is sufficiently small." However, numerical tests (which will be discussed in another report) show that the theory can be used for synchronous satellites of the Earth.

REFERENCES

1. Blitzer, L., "Satellite Resonances and Librations Associated with Tesseral Harmonics of the Geopotential," Journal of Geophysical Research, Vol. 71, (1966).

2. Brown, E. W., "Elements of theory of Resonance," Rice Inst. Pamphl. 19, (1932).

3. Cesari, L., Asymptotic Behavior and Stability Problems in Ordinary Differential Equations, New York, Academic, 1963.

4. Kaula, W. M., Theory of Satellite Geodesy, Waltham, Blaisdell, 1966.

5. Krylov, N. and Bogoljubov, N. N., An Introduction to Nonlinear Mechanics, Princeton University Press, 1947 (Annals of Mathematics Studies, vol. 2).

6. Morando, M. B., "Orbites de Résonance des Satellites de 24 H," Bulletin Astronomique, tome XXIV, Fasc. 1 (1962).

7. Morgunov, B. I., "Stationary Resonance Behavior of Certain Rotary Motions," Dokl. Akad. Nauk SSSR 155 (1964).

8. Morrison, J. A., "Application of the Method of Averaging to Planar Orbit Problems," J. SIAM 13, (1965).

9. Sansone, G., and Conti, R., Nonlinear Differential Equations, New York, Pergamon, 1952.